Understanding the Genetic Basis of Carotenoid Concentration in Lentil (*Lens culinaris* Medik.) Seeds

A Thesis Submitted to the College of

Graduate Studies and Research in

Partial Fulfillment of the

Requirements for the Degree of

Masters of Science in the

Department of Plant Sciences

University of Saskatchewan

Saskatoon

by

Tina Thomas

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the libraries of this University may make it freely available for inspection. I further agree that permission for copying this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work, or in their absence, by the Head of the Department or Dean of the College in which my thesis was done. It is understood that any copying or publication or use of the thesis, in whole or in part, for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Plant Sciences

University of Saskatchewan

51 Campus Drive

Saskatoon, Saskatchewan

S7N 5A8

ABSTRACT

Lentils are an inexpensive source of protein, vitamins and minerals. Lentil seeds contain carotenoids that have antioxidant properties and play an important nutritional role as precursors of vitamin A. Improving concentration of carotenoids in lentils has potential as component of a bio-fortification program. The understanding of the genetic control of carotenoids in lentil will help breeders develop strategies for developing varieties with higher carotenoid concentration.

The objectives of this research program were to evaluate the concentration of carotenoids in mature lentil seeds and to identify genomic regions that possibly influence carotenoid concentration. The experimental program involved:

i) analyzing the carotenoid concentration in seeds produced from the specific crosses among lentil genotypes with three cotyledon colours using high pressure liquid chromatography (HPLC)

ii) analyzing an association mapping panel to develop potential single nucleotide polymorphism (SNP) markers for genes associated with carotenoid concentration

For the first objective, dihybrid crosses were made between lentil cultivars with red, yellow and green cotyledons. Hybridized lentil populations were grown in the greenhouse and phytotron chamber up to the F_3 generation and then seeds were analyzed for carotenoid concentration. As expected, the expression of red cotyledon colour was dominant over yellow, and these two cotyledon colours were inhibited by an epistatic interaction with green cotyledon colour. Lentil seeds with green cotyledon colour had higher carotenoid concentration than red cotyledon types which in turn had higher carotenoid concentration compared to yellow cotyledon lentils.

Identifying molecular markers associated with carotenoids can be part of a crop improvement strategy for both marker-assisted selection and marker-assisted breeding (MAS; MAB). Association mapping using broad genetic materials might result in high resolution. For this purpose an association mapping panel of 143 lentil genotypes was grown at two different locations near Saskatoon, Canada, in 2011 and 2012. Concentration of three carotenoids in lentil seed samples was measured using reverse phase HPLC. Of the 143 genotypes, 60 accessions were common for both years and locations. Concentrations of lutein, zeaxanthin and violaxanthin in seed samples were determined. Genotyping was accomplished using 1536 SNP (single nucleotide polymorphism) markers of an Illumina Golden-Gate assay. It was determined

that 168 of the SNP markers were significantly associated with carotenoid concentration components using the GLM (generalized linear model) model. These putative SNPs could be used for MAS and MAB to improve selection for carotenoids in lentil to increase the nutritional value of lentil.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Albert Vandenberg for his mentorship, patience and the countless hours of discussion involving lentils, carotenoids and association mapping. I would also like to thank my committee members Dr.Bunyamin Tar'an, Dr. Michael Nickerson and Dr. Yuguang Bai for their advice and mentorship. I am thankful to the Saskatchewan Pulse Growers for providing the funding needed for my project. I would also like to extend my sincere thanks to the Pulse Crop Breeding group at the CDC which includes Hamid Khazaei, Brent Barlow, Thiago Prado, Devini DeSilva, Chandra, Marwan Diapari, Mohamed Ali, Yurdagul Ferhatoglu and countless others who provided me with technical assistance, guidance and friendship throughout my degree. My thanks also extend to the graduate students in the Department of Plant Science with whom I have been able to forge close friendships and also to my family who provided me unconditional love and support.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACN	Acetonitrile
AFLP	Amplified fragment length polymorphism
AM	Association mapping
BHT	Butylated hydroxytoluene
CDC	Crop Development Centre
DCM	Dichloromethane
FDR	False discovery rate
FAO	Food and Agriculture Organization of the United Nations
GLM	Generalized likelihood method
GMO	Genetically modified organisms
GWAS	Genome wide association mapping
HPLC	High performance liquid chromatography
ICARDA	International Center for Agriculture Research in the Dry Areas
LAM	Lentil Association Mapping
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAB	Marker assisted breeding
MAS	Marker assisted selection
MeOH	Methanol

MLM	Maximum likelihood method
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RCBD	Randomized complete block design
RIL	Recombinant inbred line
RFLP	Restriction fragment length polymorphism
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
UPGMA	Unweighted pair group method with arithmetic mean
USDA	United States Department of Agriculture
U of S	University of Saskatchewan

CHAPTER 1 Introduction and hypotheses

1.1 Introduction

Lentil (*Lens culinaris* Medik.) is a cool season legume crop. Lentil production in western Canada has been rapidly increasing since the 1970s. Lentils are a staple daily food source for many humans, as they are rich in protein, carbohydrates and micronutrients (Wang and Daun, 2004). Being rich in essential amino acids and protein, lentils form a balanced nutritional diet when consumed with carbohydrate rich cereal crops. Therefore in countries with large vegetarian populations, like India, pulse crops like lentil are heavily consumed. Lentil is the third most consumed cool season pulse crop legume after pea and chickpea in the world (FAOSTAT 2014). Among the three cotyledon colours found in lentil, consumption of red cotyledon types is more common than consumption of yellow and green cotyledon types (Singh et al., 2014). Consumption of green cotyledon lentils is not common.

Carotenoids are important nutrients for humans. They belong to the family of polyene or phytoene molecules that share a common $C_{40}H_{56}O_2$ formula (Hata et al., 2000). There are over 600 carotenoids known in nature. Carotenoids are one of the most abundant and widely distributed classes of organic pigments that naturally occur in higher plants, algae, anoxygenic photosynthetic bacteria and cyanobacteria, as well as some non-photosynthetic bacteria and fungi (Armstrong and Hearst, 1996). However, since mammals are incapable of synthesizing carotenoids, they must be obtained through diet (Qudah and Muhammad, 2009). Carotenoids are responsible for much of the colour we see in plants in nature, for example, the red colour of tomato fruits, the orange colour of carrot roots, and the fall colouration of tree leaves.

Carotenoids accumulate in nearly all types of plastids, including the chloroplasts and chromoplasts, and are therefore commonly found in most plant organs and tissues (Chaudhary and Nijhawan, 2010). In higher plants, carotenoids can accumulate in the chromoplasts of flowers and fruits resulting in bright yellow, orange and red colours that attract pollinators and agents of seed dispersal (Kato et al., 2004). Krinsky and Johnson (2005) reported that about 50 different carotenoids can be metabolized into vitamin A, and among them β-carotene has the highest provitamin A activity. However, this bioconversion is highly variable among individuals.

Due to nutritional concerns of vitamin A deficiency in humans, development of plant cultivars with increased carotenoid concentration has become a breeding objective in many crop species including maize (*Zea mays*), barley (*Hordeum vulgare*), soybean (*Glycine max*) and rice (*Oryza sativa*) (Kimura et al., 2007; Lachman et al., 2013), in order to achieve this objective genetically modified organisms (GMO) methods were formed (example golden rice). Previous studies on the carotenoid composition and content of selected grain legumes (chickpea, *Cicer arietinum*; common bean, *Phaseolus vulgaris*, and soybean) showed that β-carotene, lutein and cryptoxanthin are the major carotenoids (Siong, 1995). Lutein was found to be approximately 60% of the sum of the total carotenoids. According to the study conducted by Fernandez-Marin et al (2014) it was found that there was a change in the concentration of carotenoids, tocopherols and fatty acids in the 10 legume species as a 'side-effect' caused by the selection of other desired traits like selection for seed storage, seed dormancy and dispersal mechanism.

Little information exists regarding the concentration, biochemical profile, and heritability of carotenoid accumulation in lentil seeds. Detailed carotenoid concentrations and carotenoid profiles of lentil seeds have not been published. A baseline study on carotenoid composition and concentration in pulses was carried out at the Crop Development Centre, University of Saskatchewan. The results indicated that lutein is the carotenoid present in the highest concentration in pea (*Pisum sativum*), chickpea and lentil (Kaliyaperumal et al., 2014). Other primary carotenoids include zeaxanthin, β-carotene and violaxanthin. The present study was aimed at understanding the genetic basis of carotenoids in lentil seeds. This study will help to determine if cotyledon colour (red, yellow and green) is related to carotenoid type or concentration, providing a baseline analysis for future potential breeding objectives for the development of high carotenoid cultivars.

1.2 Research hypotheses and objectives

1.2.1 Research hypotheses

1. The concentrations of the carotenoids violaxanthin, zeaxanthin and lutein are different in lentils with red, yellow and green cotyledons under the same environmental conditions.

2. By the analysis of the phenotypic data (violaxanthin, zeaxanthin and lutein concentration) specific carotenoids associated with SNPs (single nucleotide polymorphisms) can be identified for potential use in breeding lentil genotypes with higher concentration of carotenoids.

1.2.2 Research objectives

Hypothesis 1 Objective:

Determination of the concentration of violaxanthin, zeaxanthin and lutein in green, red and yellow cotyledon lentil lines using HPLC

Hypothesis 2 Objective:

Identification of SNPs linked to carotenoid concentration in lentil seeds using association mapping.

CHAPTER 2 Literature review

2.1 Lentil

2.1.1 General description of cultivated lentil

The systematic position of the cultivated lentil (*Lens culinaris* ssp. *culinaris*) in the plant kingdom is as follows: Order-*Fabales*, Family- *Fabaceae*, Genus- *Lens*, Species- *culinaris.*

The cultivated lentil and its wild relatives are self-fertilizing diploids having chromosome number 2n=2x=14. Lentil plants grow to the height of about 15-45 cm. The leaf arrangement on the plant is alternate. Leave usually have six pairs of oblong leaflets of 15 mm length. The rachis normally develops a terminal tendril at the onset of flowering. Peduncles arise from leaf axil normally bearing two to four flowers. These flowers self-pollinate and form inflated oblong pods having one to three double convex lens shaped seeds of 2-8 mm. The cultivars of lentil differ in height, pubescence, stem colour, leaves, flower colour, pod colour, cotyledon colour, seed coat colour and seed coat pattern (Saskatchewan Pulse Growers, 2000).

There are two main classes of lentils differentiated based on seed size- *macrosperma* and *microsperma*. Cultivars with large seeds and little pigmentation are classified as *macrosperma* (seed diameter is between 6-9 mm) while those with small to moderate seed size and pigmentation belong to *microsperma* (seed diameter between 3-6 mm) (Mishra et al., 2007). Seed coat colour in lentils can range from grey to yellow, green, tan, brown or black. The seeds are lens-shaped and the colour of cotyledon colours may be red, yellow, and green or even brown. These primary external and internal seed characteristics are the basis for trade and consumption patterns of lentils around the world.

Lentil is known to have tolerance for extreme environmental conditions such as drought and high temperature, and can be grown in semi-arid regions like south-western Saskatchewan (Saskatchewan Ministry of Agriculture, 2010). Lentil crops can be grown in rotation with cereals to reduce soil erosion, to improve disease and weed control, and to reduce the need for nitrogen fertilizers in the cropping system. Lentil is capable of supplying a significant part of its nitrogen requirement by fixing nitrogen from the air through symbiotic relationship with nitrogen-fixing *Rhizobium* bacteria. Lentil has an indeterminate growth habit. In the plants flowering continues

until they encounter any forms of stress, such as heat, drought, frost, nitrogen deficiency, chemical or mechanical desiccation

2.1.2 World production and trade of lentil

Canada, India, Turkey, Australia and USA are the world's top producers of lentil accounting for well over 50% of total production. Saskatchewan cultivates the major portion of the lentil crop in Western Canada (FAOSTAT, 2014). Lentil production in Canada for the year 2014 (about 2 million Mt) was almost half the world supply (FAOSTAT, 2014). More than a million ha are now cultivated annually, and Saskatchewan is the major production region with about 95% of the lentil production in Western Canada.

Although the production of lentils and other pulses is far less than cereal and oilseed production throughout the world, production remains important because of its benefits for producers and consumers. Lentil has become an important part of the cropping system of western Canada, particularly in Saskatchewan.

2.1.3 Lentils and human nutrition

Plant seeds are a major source of nutrients for humans and animals. Adequate dietary protein is essential for human health and optimal livestock production, and pulses are one of the primary sources. Improvement in both quantity as well as quality of food is needed to cope with the increasing human population. The 'Green Revolution' in cereals was a boon to humanity as it solved the problem of starvation but fell short in addressing the health problems associated with deficiencies in vitamins and minerals.

Lentils are a part of the daily diet in many countries, as they supply both micronutrients and macronutrients and have high protein content. In western Canada, lentil protein content ranged from 25.8% to 27.1% per 100 g (Wang and Daun, 2006). In southern Asia where the major part of the population is vegetarian, lentil is an essential source of protein that provides the essential amino acids, isoleucine and lysine. Methionine and cysteine are deficient in lentil; however a sufficient level of them is present in sprouted lentils. When lentils are consumed in combination with cereals, a balanced and nutritious diet can be formulated, as cereals have a relatively high content of methionine and cysteine (Sell, 1993). This is a dietary strategy called protein combining or protein complementing.

Lentils are a good source of iron providing about 7.4mg $100g^{-1}$ and also contain dietary fiber (30g $100g^{-1}$), folates (479mcg $100g^{-1}$), vitamin B_1, and minerals (USDA). According to Bhatty (1989) calcium and iron are present in lentils in significant amounts. According to Ray et al. (2013) lentil contains significant proportions of the recommended daily allowance (RDA) of minerals for human consumption.

2.2 Carotenoids

2.2.1 Carotenoids and plants

Carotenoid colours as observed in nature ranges from red to yellow. Carotenoids can accumulate in plant cells, specifically in all types of plastids, including the chloroplasts and chromoplasts, and are thus found in most plant organs and tissues (Chaudhary and Nijhawan, 2010). In higher plants, the carotenoids can accumulate in the chromoplasts of flowers and fruits, resulting in bright yellow, orange and red colours. In these plant tissues, carotenoids serve as colorants to attract pollinators and agents of seed dispersal. Previous studies indicated that for the synthesis of the plant hormone abscisic acid epoxy-carotenoids, violaxanthin and neoxanthin are precursors (Ronen et al., 1999). About 50 different carotenoids can be metabolized into vitamin A and, and among them, β-carotene has the highest provitamin A property. However, this bioconversion is highly variable and unique for every person (Krinsky and Johnson, 2005).

Among the carotenoids, lycopene is one of the most abundant. Lycopene concentration is high in okra (*Abelmoschus esculentus*), green bean (*Phaseolus vulgaris*) and tomato (*Solanum lycopersicum*) (Qudha et al., 2009). At the grain maturity stage for durum wheat (*Triticum durum*), the predominant carotenoid is lutein (Ramachandran, 2009). In the same study zeaxanthin was reported to be the second most abundant carotenoid. Previous studies on the carotenoid composition and content of grain legume seeds (chickpea, common bean and soybean) showed that lutein, cryptoxanthin and β-carotene were the major carotenoids. Lutein concentration was highest, followed by cryptoxanthin and then β-carotene. Lutein was found to be approximately 60% of the total carotenoids (Siong, 1995).

A study by Qudah and Muhammad (2009) to determine the carotenoid composite on and concentration in six frequently consumed vegetables like okra (*Abelmoschus esculentus*), green bean (*Phaseolus vulgaris)*, eggplant (*Solanum melongena*), zucchini (*Cucurbita pepo*), carrot

(*Daucus carota* subsp. *sativus*) and tomato) indicated that the majority of carotenoids in carrot are α-carotene and β-carotene, while lutein is a minor component. Among these 6 vegetables, α-carotene exists only in carrot while β-carotene was found in all vegetables. Lycopene was found in okra, green bean, and tomato, with the highest lycopene concentration found in tomato (Dutta et al., 2005). In the case of pulses, lutein is the most accumulated carotenoid.

In saponified orange (*Citrus sinensis*) juice, thirty-nine carotenoid pigments were separated using a non-end-capped C-30 reversed phase column with solvent system of water, methanol, and methyl-tert-butyl ether gradient. Using diode array spectral characteristics, relative elution, and retention times pigments were identified when compared to the standards. At 430 nm more peaks were observed but had less selectivity for several carotenoids compared to the other two peaks at 350 nm and 486 nm (Rouseff and Raley, 1996).

2.2.2 Health benefits of carotenoids

According to Cadenas and Packer (2002) carotenoids family of natural fat-soluble nutrients found throughout the plant kingdom. They serve as photosynthetic pigments and most are known to be beneficial for human health. Many epidemiological studies have shown that carotenoids are important for normal cell regeneration (Giovannucci and Clinton, 1998; Clinton, 1999). They promote eye health and prevent premature aging (Landrum et al., 1997) in addition to numerous other health-promoting effects linked to interactions with unstable oxygen molecules called free radicals (Rao and Agarwal, 2000; Cadenas and Packer, 2002). The properties of carotenoids include their action as antioxidants, which have the ability to protect cells and tissues from the negative effects of free radicals (Paiva and Russell, 1999; Mortensen et al., 1999).

Due to the desirable health benefits of carotenoids many crop have been bio-fortified to increase the concentration of carotenoid in them a few example are golden rice, orange maize, orange sweet potato and orange cassava. A few countries like Nigeria and Zambia (orange maize, 2012), Uganda (orange sweet potato, 2007), Democratic Republic of Congo and Nigeria (orange cassava, 2011) have release them commercially for human consumption. For a an adult individual daily intake of 5-6 mg of carotenoids is essential to have the recommended concentration of carotenoids according to National Cancer institute (NCI, USA) for the uptake of Vitamin A from these pro- vitamin A compounds. According to Muller (1996) in order to have the recommended

intake of carotenoids to avoid Vitamin A deficiency, 100-200 g of vegetables and fruits (fresh sample) should be consumed daily which equals to about 5-6 servings per day. He found the concentration of carotenoids in "kale was 34.8 mg$100g^{-1}$, red peppers 27.4 mg$100g^{-1}$, parsley 25.7 mg$100g^{-1}$, spinach 17.3 mg$100g^{-1}$, lamb's lettuce 16.0 mg$100g^{-1}$, carrots 15.8 mg$100g^{-1}$ and tomatoes 12.7 mg$100g^{-1}$ headed the list of vegetables with more than 10 mg $100g^{-1}$ and for fruits "papayas 3.8 mg, grapefruits 3.6 mg, nectarines 2.9 mg and apricots 2.6 mg were pre-eminent with more than 2 mg $100g^{-1}$".

Fernandez-Marin et al. (2014) studied the carotenoid concentration in ten domesticated and wild legumes and found that among the domesticated legumes, lentils (~10.1 µg g^{-1}) has the highest carotenoid concentration followed by chickpea (~9.4 µg g^{-1}) and soybean (~6.4 µg g^{-1}), while wild legumes soybean (~27.3 µg g^{-1}) had the highest carotenoid concentration followed by lentils (~17.7 µg g^{-1}) and chickpea (17.6 µg g^{-1}) among the other ten legumes in this study. In a study conducted on selected Mediterranean legumes by El-Qudah (2014) similar results were seen with lentil (11.6 µg g^{-1}) having the highest carotenoid concentration followed by chickpea (5.6 µg g^{-1}) then faba bean (3.1 µg g^{-1}) and lastly dry bean (0.2 µg g^{-1}).

2.2.3 Structure of carotenoids

Most carotenoids are C_{40} tetraterpenoids derived from phytoene and comprise a large family of more than 700 structures (Chaudhary and Nijhawan, 2010). The most prominent chemical feature of carotenoids is the polyene chain, consisting of 3 and 15 conjugated double bonds, which are responsible for the spectral properties of carotenoid, and therefore the colour of the specific carotenoid (Ronen et al., 1999). The numbers of double bonds determine the spectral properties of a given carotenoid, which typically absorbs light between 400 and 500 nm (Armstrong and Hearst, 1996). A critical step in the formation of the C_{40} acyclic hydrocarbon is the tail-to-tail condensation of two molecules of the C_{20} intermediate geranyl- geranyl pyrophosphate (GGPP) to form phytoene (Britton, 1995; Armstrong and Hearst, 1996). The parent C_{40} carbon skeleton is the source for other individual variations. The conjugated C=C double bond system in the linear C_{40} hydrocarbon backbone is considered to be the most important factor responsible for energy transfer reactions in photosynthesis (Young, 2000). The C_{40} carbon skeleton can be modified in several ways, including cyclization at one end or both ends of the molecule to give seven different

end groups, changes in hydrogenation level and addition of oxygen-containing functional groups (Britton, 1995).

Some carotenoids have a structure with fewer or more than 40 carbon atoms. For instance, some organisms produce carotenoids containing C_{30} structures, while a few bacteria synthesize C_{45} or C_{50} carotenoids by adding isoprene to the C_{40} backbone (Armstrong and Hearst, 1996). Carotenoids that have less than 40 carbon atoms are known as apocarotenoids or norcarotenoids (Britton, 1995; Armstrong and Hearst, 1996). They differ in the method by which they lose carbon atoms from the C_{40} carotenoids. Loss of carbon atoms from the ends of C_{40} carotenoid molecule forms apocarotenoids, while the loss of carbon atoms within the chain produces norcarotenoids (Britton, 1995).

2.2.4 Carotenoid biosynthesis

The carotenoid biosynthesis pathway in higher plants (Figure 2.1) shows that the biosynthesis of carotenoids occurs within the chloroplasts of plants and algae (Cunningham et al., 1996). The first critical step is the formation of phytoene produced by the head-to-head condensation of two C_{20}-geranyl-geranyl diphosphate (GGPP) molecules under the control of phytoene synthase (Ronen et al., 1999). After the formation of phytoene, phytoene desaturase and z-carotene desaturase are the two enzymes involved in converting phytoene to lycopene (Kato et al., 2004; Clotault et al., 2008). Carotenoid isomerase is involved in the conversion of poly-cis lycopene to trans-lycopene (Kato et al., 2004). The cyclization of the ends of the linear carotenoid lycopene is an essential step in this pathway. Two enzymes, lycopene-cyclase and lycopene β-cyclase are involved within the carotenoid biosynthetic pathway to convert the linear lycopene structure to other carotenoids by the addition of rings at the end of the molecule. Lycopene β-cyclase adds a β-ring to each end of the lycopene molecule, resulting in the production of β-carotene. Lycopene ε-cyclase adds one ε-ring to lycopene resulting in the production of δ-carotene. After that, one β-ring is added to δ-carotene, which is catalyzed by lycopene β-cyclase, leading to the production of α-carotene (Cunningham et al., 1996; Clotault et al., 2008). By sequential hydroxylation, α- carotene can be converted into lutein and this reaction is catalyzed by α-ring hydroxylase and β-ring hydroxylase, and β-carotene can be hydroxylated into zeaxanthin by β-ring hydroxylase (Kato et al., 2004; Clotault et al., 2008). Furthermore, zeaxanthin is transformed into violaxanthin when it is catalyzed by zeaxanthin epoxidase. Carotenoid biosynthesis and its regulation have been studied

in various plant species, such as citrus fruit (Kato et al., 2004), tomato (Ronen et al., 1999), rice (Chaudhary and Nijhawan, 2010), carrot (Clotault et al., 2008). A study (Chaudhary and Nijhawan, 2010) on carotenoid biosynthesis genes in rice showed that 16 genes are involved in carotenoid biosynthesis. A majority of the genes are differentially expressed in different tissues, vegetative and reproductive stages, and expression of some genes is enhanced under abiotic stress conditions.

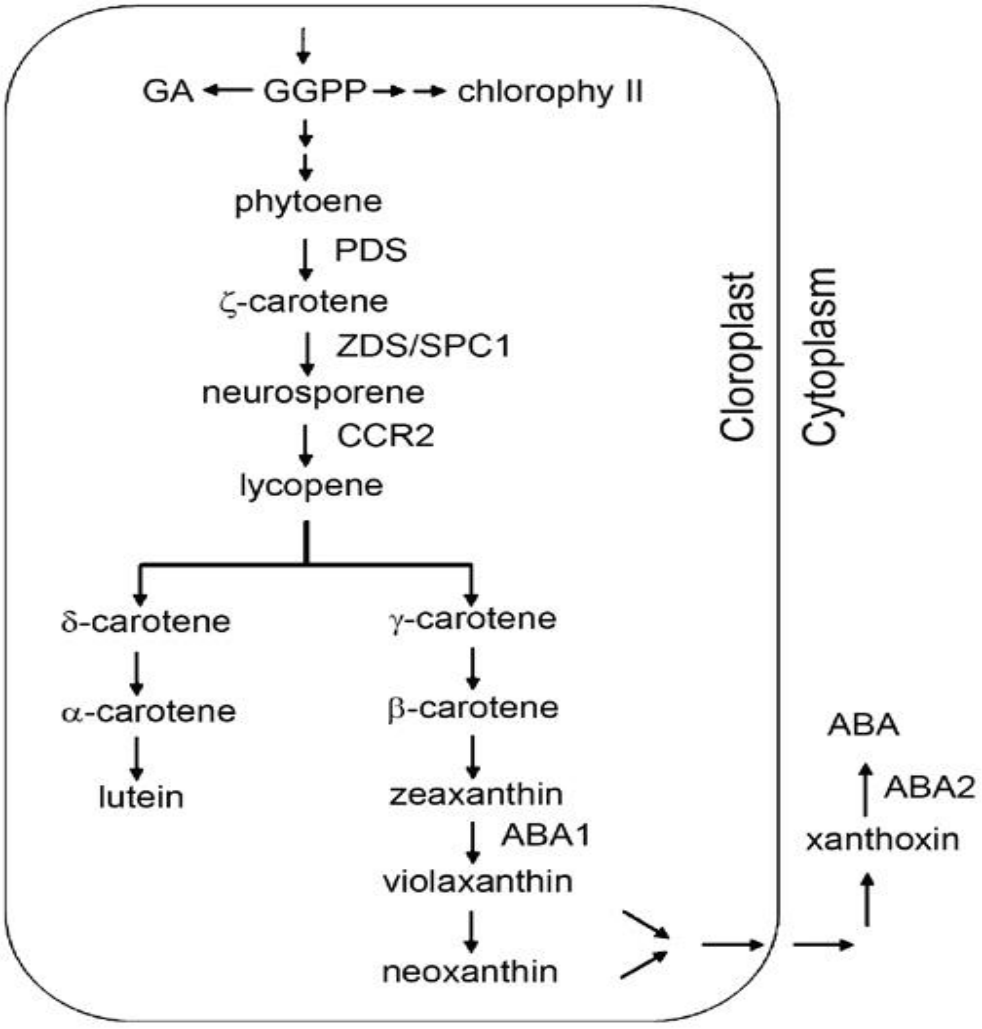


Figure.2.1 The carotenoid biosynthetic pathway (adopted from Botella- Pavia et al., 2004)

Kato et al. (2004) reviewed carotenoids and expression of carotenoid biosynthesis genes during citrus fruit maturation. They found that the concentration of total carotenoids in the juice sacs was low during the green stage of fruit development and carotenoid biosynthetic genes were expressed at various levels. After the green stage, carotenoids massively accumulated and there was

considerable variation in the concentration of different carotenoids. These two examples indicated that there are noticeable differences in expression levels of carotenoid biosynthesis genes, resulting in great variability in the concentration.

2.2.5 Functions of carotenoids

Carotenoids are responsible for most of the colours in nature. It is well known that carotenoids have the ability to absorb light. The process of light absorption involves the formation of the excited singlet state of carotenoids. This excitation energy is transferred to photo-responsive pigments such as chlorophylls to initiate the process of photosynthesis. The two major biological functions of carotenoids in photosynthetic microorganisms and plants are (1) absorption of light and transfer of energy for photosynthesis and (2) protection of chlorophyll from photo-damage (Krinsky, 1994).

Many epidemiologic studies in animal species have associated high carotenoid intake with a decrease in the incidence of chronic diseases. However, the biological mechanisms for such protection are currently unclear (Paiva and Russell, 1999). Epidemiologic studies consistently indicated that increased consumption of foods rich in ß-carotene can reduce the risk of lung and other cancers (Young, 2000). ß-carotene and other carotenoids whose structures are highly similar to ß-carotene serve as precursors for vitamin A, retinol and retinoic acid in mammals, therefore playing important roles in nutrition, vision and cellular differentiation (Armstrong and Hearst, 1996). Carotenoids are known for their antioxidant properties. Fruits and vegetables containing vitamin C, vitamin E and carotenoids including β-carotene, β-cryptoxanthin, α-carotene, lycopene, zeaxanthin and lutein have been suggested as natural sources of antioxidants. Antioxidant functions play essential roles in decreasing DNA damage, maintaining immune function and preventing the development of some diseases (Qudah and Muhammad, 2009).

The antioxidant properties of carotenoids have been suggested to be closely related to their single oxygen quenching properties and their ability to trap peroxyl radicals (Stahl and Sies, 1996). The ability to quench single oxygen enables carotenoids to maintain an excited state. This ability through a series of rotational and vibrational interactions with the solvent helps dissipate newly acquired energy, thus regenerating the original unexcited carotenoid. The regenerated unexcited carotenoid can be reused for further cycles of singlet oxygen quenching (Paiva and Russell,

1999). The ability to quench single oxygen primarily depends on the structure of carotenoid molecule. It is the conjugated C=C double bond system in carotenoid structure that allows the quenching of single oxygen molecule (Young, 2000). Another form of β carotene like ί-carotene and other carotenoids can function as scavengers of peroxyl radicals (Rice-Evans et al., 1997). Carotenoids are considered to interact with peroxyl radicals via an unstable ί-carotene radical adduct. Carotenoid adduct radicals are highly resonance-stabilized and are predicted to be relatively unreactive. The adduct radicals may be degraded to produce non-radical products, and they may terminate radical reactions by binding to the attacking free radicals.

2.2.6 Genetic studies of carotenoids

Genetic studies on cotyledon colour in lentil are of academic interest and economic importance. Globally, lentils with orange (also known as red) and yellow cotyledons are most commonly consumed. Green cotyledon lentils are also becoming available. Lentils with red cotyledons were traditionally consumed in South Asia and parts of the Middle East. Yellow cotyledon lentils were traditionally consumed in most of the Mediterranean region and Central Asia. Previous genetic studies indicated that orange cotyledon (*Yc*) is dominant to yellow cotyledon (*yc*) and both orange and yellow cotyledon had epistatic interaction with the gene (*I-yc*) which causes cotyledons to be green when it is in the double recessive state (*i-yc i-yc*) (Slinkard, 1978).

Emami and Sharma (1996) reported a more elaborate genetic model for cotyledon colour of lentil. Two types of yellow in lentil cotyledon were reported; bright yellow and dull yellow with a brownish tinge. The latter one was called brown to distinguish it from bright yellow. They reported that pigment synthesis in lentil cotyledons was controlled by two genes: *Y* and *B* (*Y*-yellow, *B*-brown). The *Y_bb* genetic configuration determines synthesis of the bright yellow pigment while *yy_B* results in brown cotyledon. Orange colour is due to the interaction of two dominant genes. The orange colour will be produced only when both the genes are present in dominant condition (*Y_B_*). The absence of both *Y* and *B* in the double recessive state (*yybb*) results in light green cotyledons.

A third gene (*Dg*) which results in the production of dark green cotyledons was reported by Emami and Sharma (2002). This gene behaved as a monogenic recessive to the orange phenotype. The cotyledon colour in lentil is controlled by a system of three genes (*Dg*-dark green). In the

presence of dominant gene *Dg*, the yellow cotyledon is produced by the gene *YY* while the gene *BB* produces brown (brownish yellow) cotyledon. When all the genes are present in dominant conditions (*Dg_Y_B_)*, the orange cotyledons are produced. When the gene for dark green colour is recessive (*dgdg*), irrespective of dominant or recessive genes for yellow/ brown/ orange cotyledons (*YY* or *yy*; *BB* or *bb*; *Y-B-*), the dark green cotyledons are developed. The light green cotyledons are produced when the genes for yellow and brown colours are both recessive (*Dg_yybb*). This research has never been independently confirmed.

2.3 Association mapping

2.3.1 Molecular markers

Molecular markers are the tags that help identify the genomic region of interest and can be valuable as a plant breeding technique as an alternative or supplement to phenotypic selection. Depending on the linkage between the marker and the trait of interest, they can be inherited together or in association with the trait. This would help in coming to phenotypic conclusions for an individual. Quantitative trait loci (QTL) are fragments of DNA that are associated or linked with genomic regions that have genes/ alleles which express a quantitative trait. To determine an association in addition to the use of morphological markers in plant breeding, various molecular marker techniques were developed, including isozymes, restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), microsatellites or simple sequence repeat (SSR), and single nucleotide polymorphism (SNP) (Jiang et al., 2013).

Association mapping has become an important tool for helping scientists to explore the individual genes associated with quantitative traits by using both molecular marker and phenotypic trait data (Miles et al., 2008). Association analysis is a technique used to develop understanding of the genetic basis of variation in complex traits by statistically analyzing genotypic and phenotypic data (Young et al., 2000). Association mapping is also known as linkage disequilibrium mapping, a method of QTL detection consisting of identifying marker-trait associations in genetically diverse populations.

In agricultural applications in plant science and plant breeding, association mapping can be used to design crosses to improve products (Breseghello, 2006). DNA markers that are tightly linked to

a QTL can be designed as molecular tools for marker-assisted plant breeding (MAB) (Ribaut and Hoisington, 1998). The mapping population can be divided by the application of markers into diverse genotypic groups according to the presence or absence of a particular marker locus (Collard et al., 2005).

Among the marker used in plant breeding SNPs have been the most suitable owing the fact that they are the most cost effective in terms of labour and time, and highly reproducible with wide genome coverage. They are useful for studies involving cultivar discrimination. RFLP was one of the initial techniques employed for molecular studies but because they need large quantity of good quality DNA along with radioactive labeling they are expensive and labour intensive. In comparison AFLPs and RAPDs may be less costly but have poor reproducibility and the results are inconsistent because patterns may change depending on the genetic material used. SSRs are highly reproducible and have greater genome coverage but due to the inability to use them across species and demand of sequencing data they are not as advantageous as SNPs (Kumar et al., 2009).

2.3.2 Population structure and kinship

Population structure is a method used to determine if the population groups or subgroups have deviated from the Hardy-Weinberg equilibrium. Having an elevated structure can cause spurious association in mapping populations. Another factor that may cause high association rate in mapping analysis is kinship, which is the coefficient of relatedness between two individual chosen at random (Sajjad 2012).

2.3.3 Linkage disequilibrium (LD) and association mapping

The completely independent inheritance of two genes is referred to as linkage disequilibrium. If two genes are in linkage disequilibrium it could be due to a genetic linkage. Genetic linkage is the state in which genes are located closely on the same chromosome.

Association mapping or linkage disequilibrium (LD) mapping is the technique of linking phenotype to genotype using the ancestral LD to map the QTLs. According to Oraguzie et al. (2007), LD is the ‘non- random co-segregation of alleles at different loci’. Identifying individual genes for crop improvement using marker assisted selection can be time consuming and expensive. Association mapping is advantageous over QTL mapping because it identifies

significant associations only for those markers that are closely linked to the genomic region of interest. There are two approaches in association mapping: genome–wide association mapping (GWAM) and candidate gene association mapping (CGAM) (Pritchard et al., 2000).

GWAM involves the identification of significant associations in the entire genome of the species. This technique requires a high density of markers to cover the entire genome depending on the genome size and LD decays. The CGAM method is more specific as it only focuses on the candidate gene-marker association instead of a genome wide scan. Prior knowledge about the candidate gene and the phenotype are required, which are provided by the QTL mapping data or the GWAM data.

CHAPTER 3

Study 1: Carotenoid variability and concentration in relation to cotyledon colour of lentil in segregating populations.

Hypothesis: The concentrations of carotenoids- violaxanthin, zeaxanthin and lutein are different in lentils with red, yellow and green cotyledons under the same environmental conditions.

Objective: Determination of the concentration of violaxanthin, zeaxanthin and lutein in green, red and yellow cotyledon lentil lines using HPLC.

3.1 Background

Morphological colour and pattern characteristics of seed coats and cotyledons of lentils are important commercial attributes and many of these characteristics are inherited as single genes. Some characteristics are influenced by environmental factors and/ or genetic factor. Seed cotyledon colour is genetically determined and is an important cost-efficient, simple and rapid morphological marker for identifying hybrid seeds and in the subsequent segregating population depending on the parents involved in the cross (Erskine, 2009).

According to Slinkard (1978) expression of the orange/ red cotyledon colour phenotype is the result of a single gene by which the *Yc* allele is dominant over *yc*, which expresses yellow cotyledon. Both alleles can only be expressed when an inhibitor gene (*I-yc*) is present in dominant form a second unlinked locus. The gene designation *I-yc* where 'I' stands for inhibitor is used because in the double recessive state, *i-yc i-yc*, the phenotypic cotyledon colours red and yellow are not expressed.

In a study conducted to determine the influence of cotyledon colour on carotenoid concentration by Kaliyaperumal et al. (2014) it was identified that green cotyledon peas had higher carotenoid concentration than yellow cotyledon peas.

3.2 Materials and Methods

This study of lentil was designed to determine the concentration of three carotenoids- violaxanthin, lutein and zeaxanthin in lentil seed tissue of segregating populations derived from

crosses between genotypes that express yellow (CDC Greenstar), red (CDC Maxim) and green (CDC QG- 2) cotyledon phenotypes. Carotenoid concentration was analyzed in mature lentil seeds produced from plants of parents and segregating generations of specific crosses between lentil genotypes with red, green and yellow cotyledon colours. Plants were grown in the greenhouse and phytotron chambers at the University of Saskatchewan (U of S). Standard curves were generated for the three carotenoids of interest. This type of analysis was used to quantify these three individual carotenoids in the lentil seeds from the segregating generations. The sum of the three individual carotenoids concentrations gave the total concentration.

3.2.1 Plant material

The three selected genotypes were all well- adapted commercial lentil cultivars developed at the Crop Development Centre (CDC) at the U of S (Table 3.1), representing the three classes of cotyledon colour.

Table 3.1 Lentil genotypes for developing populations for carotenoid analysis

Genotype	Seed size	Seed coat colour/ pattern	Cotyledon colour
CDC Greenstar	Large	Green, no pattern	Yellow
CDC Maxim	Small	Grey, no pattern	Red
CDC QG- 2	Small	Green, marbled pattern	Green

Parental lines (Table 3.1) were grown in the phytotron and greenhouse facilities at the U of S. The growth conditions for a growth chamber (Conviron, Winnipeg, MB, Canada) located at the College of Agriculture and Bioresources, U of S were set at 22°C/ 18°C (day/ night) and photoperiod was maintained at 16 h daylight and 8 h darkness. Seeds were planted in 1 gallon pots filled with a 50/50 (v/v) mixture of soil media Sunshine Mix No. 3 and No. 4 (Sun Grow Horticulture, Vancouver, BC, Canada) in January 2012. The seeds were scarified before planting using a sharp blade, and upon germination pots were watered and fertilized thrice and once a

week, respectively. Fertilizer solution of N, P and K (20:20:20, Plant Products Co. Ltd, ON, Canada) was added to pots once a week, during the entire vegetative growth period of the plant. Flowering fertilizer (15:30:15, Plant Products Co. Ltd, ON, Canada) was added during the flowering stage. Once the plants reached the flowering stage, specific crosses between the cultivars of were made between the plants which were as follow, CDC Greenstar × CDC QG- 2, CDC Maxim × CDC QG- 2 and CDC Maxim × CDC Greenstar. Plants were harvested at maturity in mid to late March, 2012. The F_1 seeds from these crosses were planted again in the growth chambers and grown under similar conditions beginning in late March 2012.

The cotyledon colour of each F_1 seed was noted before planting. In order to increase the population size of the F_1 generation, to maximize production of F_2 seeds, cuttings from the lower nodes were planted in cell trays (32- cell deep 12cm Root Trainers) filled with a mixture of soil medium Sunshine Mix No. 3 and No. 4 (Sun Grow Horticulture, Vancouver, BC, Canada) once the F_1 plants reached the 5-branch stage. Prior to planting, cuttings were lightly dipped in Stim-Root Rooting Powder (0.1% IBN, Plant Products Co. Ltd, ON, Canada). About 7- 10 nodes cutting were taken from each cross. Plantlets derived from cuttings were transplanted into 1 gallon pots after roots were formed. The F_1 plants were raised in the same manner as their parent plants, and the F_2 seeds were harvested at maturity.

The F_2 seeds were separated based on the cotyledon colour by scarifying the seeds which were then planted again in 1 gallon pots. The cotyledon colour of each F_2 seed was noted before planting. For each cross 125- 130 F_2 seeds were planted at the U of S, Department of Plant Science greenhouse facility under same environmental conditions reported earlier. The F_3 seeds from each F_2 plant was harvested at maturity and scarified to determine cotyledon colour. The segregation ratio for cotyledon colour in both generations F_2 and F_3 was determined from cotyledon colour segregation by comparison with genetic ratio expectations (Table 3.2).

HPLC analysis was used to determine carotenoid concentrations for parents, F_2 and F_3 generation seeds from the three crosses. Ten F_2 seeds were randomly selected from all seeds produced by their respective F_1 parent plants irrespective of their cotyledon colour and were planted to develop the F_3 generation. HPLC analysis is a destructive analysis hence the same seed cannot be analysed and planted to develop the next generation. The rest of the F_2 seeds from each cross were separated based on cotyledon colour and analysed for carotenoid concentration after separating

them based on cotyledon colour within each cross. Likewise the F_3 seeds were separated into cotyledon colour from each F_2 plant and a subsample was analyzed by cotyledon colour grouping. The same procedure was repeated for the other two crosses.

The number of yellow, red and green cotyledon seeds produced from each cross was recorded for all three generations. The cotyledon colour of the F_2 seeds and the proportions of F_3 seeds phenotypes produced from the F_2 plants along with the number of seeds are shown in Tables 3.2 for each cross. For each sample rep two HPLC technical replications were conducted.

The seed samples were separated by cotyledon colour and then ground using a Udy mill (Cyclone Sample Mill, UDY Corporation, Fort Collins, Colorado, USA) for HPLC analysis. For analysis of the F_3 generation seeds of 10 F_2 plants (~10-15 seeds each plant) was separated based on cotyledon colour and were analysed by HPLC in the same way as the F_2 seeds.

3.2.2 Standard Calibration Curve for Carotenoids

A linear standard curve was developed for analysis of the carotenoids involved in this study. About 2 ng of lutein, violaxanthin and zeaxanthin standards were injected after mixing them separately in the extraction solvent dichloromethane (DCM) and MeOH v/v (1:1) and 0.1% butylated hydroxytoluene (BHT). The stock solutions for the standards were stored at -80ºC. The standard chromatographic peaks were identified by comparing the retention time with absorption spectra. The carotenoids were detected at 450 nm wavelength (Rouseff and Raley, 1996).

3.2.3 Carotenoid Analysis

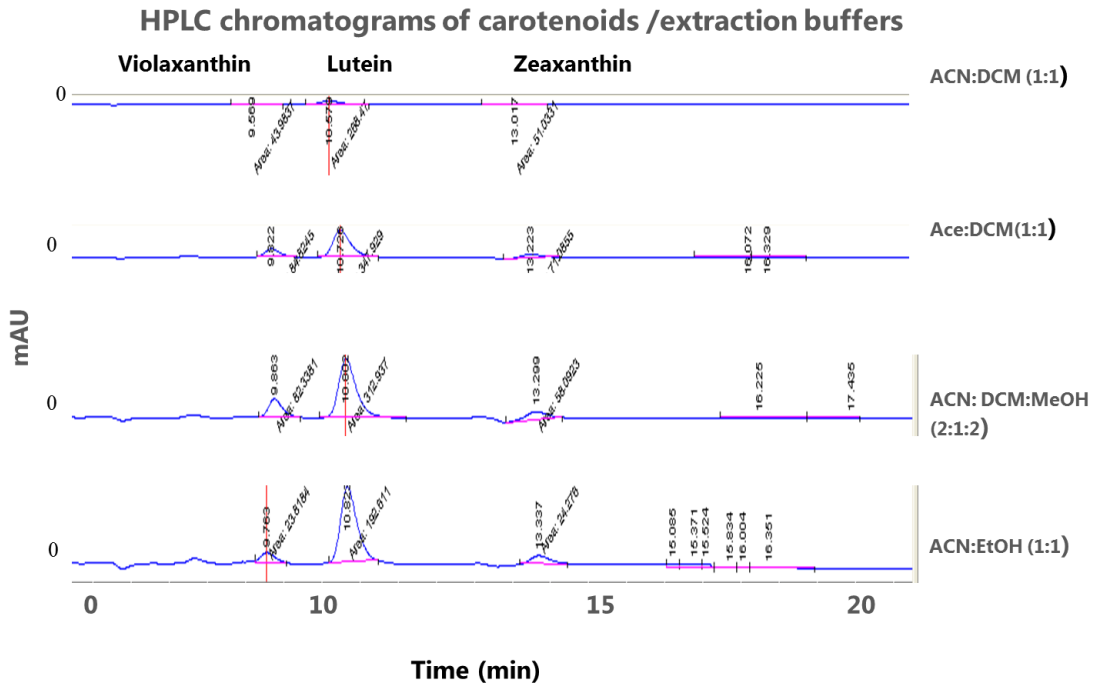
Analysis for carotenoid, violaxanthin, lutein and zeaxanthin was performed on ground lentil seeds for the three generations (parents, F_2, and F_3), beta carotenoid was not detected by the HPLC, and hence the study involved only the carotenoids- violaxanthin, lutein and zeaxanthin. The harvested F_3 seeds were dehulled using a Satake TMO5 abrasive dehulling mill (Satake USA Inc., Stafford, Texas) ground to the size of 0.3 mm using a UDY Cyclone Sample Mill (UDY Corporation, Fort Collins, Colorado). Between each grinding, the UDY mill was properly cleaned. The ground samples were stored in zipper seal plastic bags (16.5 cm × 8.25 cm, Great Value$^{TM/MC}$) to prevent dampening of the samples. The plastic bags were placed inside brown paper envelopes to prevent the carotenoids in the samples from being degraded by exposure to light and were stored at room temperature.

Since carotenoids are non-polar compounds, reverse phase HPLC was used for analysis. An Agilent 1200 HPLC system (Agilent Technologies, Palo Alto, CA) was used for this study. The system had a quaternary pump with inline degasser, auto sampler, thermostatic column compartment, and a diode array detector (DAD).

Approximately 0.1 g of each ground sample was placed into a 2 ml Eppendorf tube (1.7 ml micro-centrifuge tubes, VWR International, US) with 400 µl (1:1 + 0.1 BTH) methanol: DCM + BHT (butylated hydroxytoluene) to extract carotenoids. For each sample, three biological replicates and two technical replicates were analysed. Carotenoids were extracted by adding 400 µl DCM and acetonitrile solvent in the ratio 1:1 respectively. To prevent carotenoid breakdown in solution, 0.1% BHT was added to the solution vortexed and then centrifuged for 15 min at 11,000 rpm. The supernatant was decanted into a 2 ml Eppendorf tube and equal volume of 100% ACN + 0.1% BHT was added. These were centrifuged again at 11,000 rpm for 5 min. The supernatant was then filtered through a disc filter (13 mm syringe filter, Fisher brand, Ireland) mounted on a 1 ml syringe (BD Syringes, US) into the inserts in the amber glass vials and capped (12 × 32 mm Amber Target, DP I-D Kit, National Scientific, USA) for HPLC analysis (detailed formula for carotenoids concentration from the HPLC graph peak values is presented in Appendices 5).

For chromatographic separation of carotenoids we used- a Prodigy 5 µm ODS3100A (250 × 4.60 mm) column (Phenomenex, Torrance, CA) with the mobile phase 58:20:22 acetonitrile (ACN)/ dichloromethane (DCM)/ methanol (MeOH) flowing at 0.8 mL min^{-1}. The injection volume for each sample was 100 µl. The total run time for each sample was 30 min. Compound detection was achieved using a photodiode array detector monitoring at a 450 nm wavelength. The data was collected and analysed, then the types and amount of each of the carotenoids was tabulated.

Figure 3.1 Example of carotenoid peaks produced for CDC Redberry samples using four different extraction buffers showing peaks for the three carotenoids violaxanthin, lutein and zeaxanthin



mAU- micro-absorbance units

3.2.4 Statistical Analysis

For each cross the X^2 test was done to confirm that the crosses followed Mendelian segregation pattern in the F_2 generation. The concentrations of the three carotenoid components were converted to µg g^{-1}. There were two technical repeats per biological replicate for each plant sample for parents, F_2 and F_3 generations. ANOVA (analysis of variance) was performed using SAS 9.3 (SAS Institute Inc., Cary, NC, USA).

3.3 Results

F_1 plants of the yellow × green cotyledon cross produced yellow cotyledon seeds. For both red × green and red × yellow crosses F_1 seeds had red cotyledon colour. Table 3.1 shows that for all crosses of the three cotyledon colour combinations, the F_2 segregation ratios conformed to the

expected phenotypic and genotypic patterns based on the two gene models of the *yc* and *i-yc* genes and their interaction (Table 3.2).

Table 3.2 Observed offspring segregation ratio, value of X^2 test for red, yellow and green cotyledon colour among F_1- F_2 generations (cross 6116, CDC Greenstar× CDC QG- 2; cross 6124,CDC Maxim × CDC QG- 2; cross 6126, CDC Maxim× CDC Greenstar)

Cross	F_1	F_2 population					
		Red	Yellow	Green	X^2	*P*-value	Segregation ratio
6116	Yellow	-	182	72	1.52	0.218	3:1
6124	Red	116	24	38	6.10	0.047	9:3:4
6126	Red	151	65	-	2.99	0.084	3:1

3.3.1 Cross- 6116 (CDC QG- 2 × CDC Greenstar)

For cross 6116, between green × yellow cotyledon lentil, F_1 cotyledons were expected to be 100% yellow which is dominant over green according to Slinkard (1978), and the F_2 would give the phenotypic ratio 3 yellow: 1 green (Table 3.3).

Table 3.3 Parental lines, F_1 and F_2 phenotypes, genotypes and expected ratios for cotyledon colour for cross 6116 between yellow cotyledon CDC Greenstar × green cotyledon CDC QG-2

Generation	Genotype	Cotyledon colour phenotype	Cotyledon colour genotype at *i-yc* and *yc* loci	Expected proportion
Parent 1	CDC QG- 2	Green	*i-yc i-yc yc yc*	100%
Parent 2	CDC Greenstar	Yellow	*I-yc I-yc yc yc*	100%
F_1	(Hybrid seeds)	Yellow	*i-yc I-yc yc yc*	100%
F_2	Selfed F_1	Green	*i-yc i-yc yc yc*	25%
		Yellow	*i-yc I-yc yc yc*	50%
		Yellow	*I-yc I-yc yc yc*	25%

Table 3.4 Mean concentration of violaxanthin, lutein and zeaxanthin in lentil seeds at the F_2 and F_3 generations for cross 6116 between CDC Greenstar (Y) × CDC QG- 2 (G)

		F_2			F_3		
Genotype	**Cotyledon Colour**	**Vio (µg g^{-1})**	**Lut (µg g^{-1})**	**Zea (µg g^{-1})**	**Vio (µg g^{-1})**	**Lut (µg g^{-1})**	**Zea (µg g^{-1})**
6116- 16	Y	0.77	7.37	0.3	0.99	9.44	0.32
6116- 20	Y	1.07	6.74	0.24	1.08	9.36	0.31
6116- 26	Y	0.86	6.77	0.35	1.02	9.64	0.32
6116- 55	Y	1.17	11.75	0.26	1.12	11.75	0.29
6116- 60	Y	1.1	8.24	0.21	0.9	8.98	0.32
6116- 66	Y	0.9	7.52	0.26	1.03	9.02	0.3
6116- 84	Y	0.84	5.89	0.29	1.13	9.2	0.34
6116- 123	Y	1.15	9.72	0.28	1.24	11.9	0.24
6116- 4	G	0.84	6.77	0.25	1.6	7.06	0.54
6116- 95	G	1	7.49	0.22	0.95	9.39	0.31
Mean yellow	Y	0.92	7.1	0.27	0.99	6.9	0.41
Mean green	G	1.16	10.74	0.27	1.14	11.73	0.25
LSD $_{0.05}$	-	0.22	0.34	0.04	0.23	4.47	0.15
CDC QG- 2	G	1.38	10.99	0.43	-	-	-
CDC Greenstar	Y	1.11	7.64	0.24	-	-	-
LSD $_{0.05}$	**-**	**0.11**	**0.18**	**0.33**	-	-	-

Vio- violaxanthin; Lut- lutein; Zea-zeaxanthin, Y-yellow; G-green, LSD- Least Significant Difference

For all the genotypes in all three generations in the cross 6116 lutein showed the highest concentration followed by violaxanthin and then zeaxanthin. For cross 6116, the genotype 6116-55 had the highest concentration of violaxanthin and lutein, and was significantly different from others in F_2, but similar to the parents for violaxanthin and to the green parent for lutein. For the zeaxanthin the highest concentration in the genotype 6116-26 which was also significantly different from other in the same generation (Table 3.4).

In F_3 genotype 6116-4 had the highest concentration for violaxanthin and zeaxanthin and was significantly different from the other genotypes of the F_3 for this trait. Genotype 6116-123 showed the highest concentration for lutein and was significantly different from other genotypes in F_3.

Lutein and violaxanthin mean concentrations were higher in green cotyledon compared to yellow for both the generation F_2 and F_3. In F_2 generation the mean concentration of zeaxanthin for both yellow and green cotyledon seeds were the same, while in the F_3 generation the yellow had higher mean zeaxanthin concentration than green cotyledon seeds (Table 3.4).

Table 3.5 Mean and variance components of carotenoid concentration for three generations for cross 6116 between CDC Greenstar (Y) × CDC QG- 2 (G)

Cross	Trait	Generation	Mean±SD (μg g^{-1})	CV %	Range	F-values and Pr>F for genotype	Variance Components	
							σ^2g	σ^2e
6116	Vio	Parents	1.25±0.19	8.94	1.11-1.38	5.98	0.03	0.01
		F_2	0.97±0.16	9.96	0.77-1.17	4.53*	0.02	0.01
		F_3	1.11±0.20	12.8	0.90-1.60	5.10***	0.01	0.03
	Lut	Parents	9.31±2.37	0.15	7.64-10.99	58159.20**	5.59	0.01
		F_2	7.83±1.72	1.92	5.89-11.75	263.74***	2.96	0.02
		F_3	9.58±1.39	28.33	7.06-11.90	0.56	0.00	7.32
	Zea	Parents	0.33±0.13	7.71	0.24-0.43	51.58	0.02	4×10^{-4}
		F_2	0.27±0.04	7.16	0.21-0.35	9.63**	16×10^{-4}	3×10^{-4}
		F_3	0.33±0.08	28.04	0.24-0.54	1.48	0.00	0.01

*, ** and *** Significant at the 0.05, 0.01and 0.001 probability levels respectively; 6116- CDC Greenstar (Y) × CDC QG- 2 (G), Vio- violaxanthin, Lut- lutein, Zea-zeaxanthin.

The ANOVA indicated that for cross 6116, between yellow and green parents, there was no significant difference for violaxanthin and zeaxanthin concentration but for lutein there was significant difference (P<0.05). F_2 generation data showed significant differences among genotypes for all three carotenoid concentrations. This was not the case with the F_3 generation where zeaxanthin concentration was not significantly different among genotypes at P<0.05 (Table 3.5).

3.3.2 Cross- 6124 (CDC Maxim × CDC QG- 2)

For the cross 6124 between green and red cotyledon lentil, the F_1 seeds were expected to be 100% red since red is dominant over yellow at the *Yc* locus except when it is inhibited by the epistasis of the recessive gene *i-yc* allele according to Slinkard (1978). The F_2 for green x red crosses should segregate 9:3:4 (Table 3.6).

Table 3.6 Parental lines, F_1 and F_2 phenotypes, genotypes and expected ratios for cotyledon colour for cross 6124 between red cotyledon CDC Maxim × green cotyledon CDC QG- 2

Generation	Genotype	Cotyledon colour phenotype	Cotyledon colour genotype at *i-yc* and *yc* loci	Expected proportion
Parent 1	CDC Maxim	Red	*I-yc I-yc Yc Yc*	100%
Parent 2	CDC QG- 2	Green	*i-yc i-yc yc yc*	100%
F_1	(Hybrid seeds)	Red	*i-yc I-yc Yc yc*	100%
F_2	Selfed F_1	Yellow	*I-yc i-yc yc yc / i-yc I-yc yc yc/ I-yc I-yc yc yc*	18.75%
		Red	*I-yc i-yc yc yc / I-yc I-yc YC yc /I-yc i-yc YC YC / I-yc i-yc YC yc*	56.25%
		Green	*I-yc i-yc yc yc / i-yc i-yc YC yc/ i-yc i-yc yc yc*	25%

In the cross 6124 lutein had the highest concentration among the three carotenoids among all the genotypes in all three generations followed by violaxanthin and then zeaxanthin. For cross 6124

the F_2 genotype 6124-44 had the highest concentration of lutein and violaxanthin. Concentrations of violaxanthin and lutein and were similar to both the parents in the case of violaxanthin and to the red parent for lutein. For zeaxanthin the genotype 6124-30 had the highest concentration and was similar to both the parents in F_2 (Table 3.7).

Table 3.7 Mean concentrations of violaxanthin, lutein and zeaxanthin in lentil seeds at the F_2 and F_3 generations for cross 6124 between CDC Maxim (R) × CDC QG- 2 (G)

Genotype	Cotyledon Colour	F_2			F_3		
		Vio (µg g^{-1})	Lut (µg g^{-1})	Zea (µg g^{-1})	Vio (µg g^{-1})	Lut (µg g^{-1})	Zea (µg g^{-1})
6124- 5	R	1.09	10.6	0.37	1.09	12.79	0.49
6124- 10	R	1.15	11.69	0.49	1.42	12.79	0.35
6124- 14	R	0.97	12.79	0.47	1.37	12.38	0.44
6124- 25	R	1.17	11.94	0.45	1.11	12.67	0.35
6124- 30	R	0.98	11.94	0.56	1.19	12.18	0.43
6124- 70	R	1.01	12.67	0.36	1.25	12.28	0.52
6124- 93	Y	1.03	12.61	0.41	1.24	12.18	0.46
6124- 113	Y	1.12	11.78	0.37	1.1	12.94	0.45
6124- 19	G	1.34	11.99	0.53	1.47	13.02	0.42
6124- 44	G	1.59	13.84	0.49	1.54	13.13	0.52
Mean red	R	1.04	12.09	0.44	1.08	12.49	0.47
Mean yellow	Y	1.3	11.73	0.43	1.18	11.94	0.46
Mean green	G	1.46	12.91	0.51	1.42	13.09	0.39
LSD 0.05	-	0.2	0.22	0.04	0.28	0.89	0.07
CDC Maxim	R	1.25	8.44	0.36	-	-	-
CDC QG- 2	G	1.38	10.99	0.43	-	-	-
LSD 0.05		0.39	0.11	0.42	-	-	-

Vio- violaxanthin; Lut- lutein; Zea-zeaxanthin, Y- yellow, R- red, G- green

For lutein and violaxanthin, genotype 6124-44 had the highest concentration. The concentration of violaxanthin was close in both parents. For lutein the concentration was close to only the red parent for both F_2 and F_3 generations. Genotypes 6124-44 also showed higher concentration along with 6124-70 for zeaxanthin. At F_3, 6124-44 had the highest concentration for total carotenoids (Table 3.7).

Among the three cotyledon colours at the F_2 generation, green cotyledon seeds had the highest mean concentration for all three carotenoids followed by red and then yellow. Especially lutein concentration was highest for all genotypes and cotyledon colour followed by violaxanthin and then zeaxanthin. At the F_3 generation a similar trend was observed for the carotenoids violaxanthin and lutein but for zeaxanthin, red cotyledon lentils had the highest mean concentration followed by yellow and green cotyledon had the lowest concentration (Table 3.7)

Table 3.8 Mean and variance components of carotenoid concentration for the three generation in cross 6124 between CDC Maxim (R) × CDC QG- 2 (G)

Cross	Trait	Generation	Mean±SD (μg g^{-1})	CV%	Range	F-values and Pr>F for genotype	Variance Components	
							σ^2g	σ^2e
6124	Vio	Parents	1.32±0.09	2.33	1.25-1.38	18.62	6×10^{-3}	5×10^{-3}
		F_2	1.14±0.19	7.62	0.97-1.59	9.69**	0.03	8×10^{-3}
		F_3	1.28±0.16	15.13	1.09-1.54	2.47*	7×10^{-3}	0.04
	Lut	Parents	9.71±1.80	0.09	8.44-10.99	82787.90**	3.24	8×10^{-3}
		F_2	12.18±0.86	0.79	10.60-13.84	156.32***	0.73	9×10^{-3}
		F_3	12.64±0.36	4.85	12.18-13.13	1.36	0.02	0.37
	Zea	Parents	0.39±0.05	8.49	0.36-0.43	4.66	2×10^{-3}	6×10^{-4}
		F_2	0.45±0.07	3.68	0.36-0.56	37.20***	5×10^{-3}	4×10^{-4}
		F_3	0.44±0.06	11.35	0.35-0.52	6.26***	3×10^{-3}	2×10^{-3}

*, ** and *** Significant at the 0.05, 0.01 and 0.001 probability levels respectively; 6124- CDC QG- 2 (G) × CDC Maxim (R). Vio- violaxanthin; Lut- lutein; Zea-zeaxanthin

For the violaxanthin concentration in the red × green cross 6124, there was no significant difference between the parents, while the F_2 and F_3 genotypes had significant differences (P<0.05). For lutein the parental genotypes and the F_2 had significant differences among the genotypes, while for the F_3 there was no significant difference for this trait at P<0.05. Similar to the violaxanthin there was no significant difference among the parents and high significant difference among the genotypes of F_2 and F_3 at P<0.05 for zeaxanthin (Table 3.8).

3.3.3 Cross- 6126 (CDC Maxim × CDC Greenstar)

For the cross 6126 between parents with red and yellow cotyledons the F_1 seeds were expected to be 100% red since red is dominant to yellow (Slinkard, 1978), and the F_2 would give a phenotypic ratio of 3 red: 1 yellow cotyledon (Table 3.9).

Table 3.9 Parental lines, F_1 and F_2 phenotypes, genotypes and expected ratios for cotyledon colour of lentil for cross 6126 between red cotyledon CDC Maxim × yellow cotyledon CDC Greenstar

Generation	Genotype	Cotyledon colour phenotype	Cotyledon colour genotypes at *i-yc* and *yc* loci		Expected proportion
Parent 1	CDC Maxim	Red	*I-yc I-yc*	*Yc Yc*	100%
Parent 2	CDC Greenstar	Yellow	*I-yc I-yc*	*yc yc*	100%
F_1	(hybrid seed)	Red	*I-yc I-yc*	*Yc yc*	100%
F_2	Selfed F_1	Red	*I-yc I-yc*	*Yc Yc*	25%
		Red	*I-yc I-yc*	*Yc yc*	50%
		Yellow	*I-yc I-yc*	*yc yc*	25%

In the F_2 generation red cotyledon seeds had higher mean concentration for both violaxanthin and lutein over yellow cotyledon seeds but for zeaxanthin yellow cotyledons had higher mean concentration than in red cotyledon seeds. In the F_3 generation the mean concentration for zeaxanthin and lutein was higher in red cotyledon seeds than yellow cotyledon seeds but for

violaxanthin yellow cotyledon seeds had higher mean concentration than red cotyledon seeds (Table 3.10).

Table 3.10 Mean concentration of violaxanthin, lutein and zeaxanthin in lentil seeds at the F_2 and F_3 generations for cross 6126 between CDC Maxim (R) × CDC Greenstar (Y)

Genotype	Cotyledon Colour	F_2			F_3		
		Vio ($\mu g\ g^{-1}$)	Lut ($\mu g\ g^{-1}$)	Zea ($\mu g\ g^{-1}$)	Vio ($\mu g\ g^{-1}$)	Lut ($\mu g\ g^{-1}$)	Zea ($\mu g\ g^{-1}$)
6126- 27	Y	1.01	11.54	0.38	1.15	12.88	0.25
6126- 108	Y	1.16	10.31	0.36	1.27	9.19	0.33
6126- 110	Y	1.32	9.84	0.36	1.18	8.83	0.28
6126- 5	R	1.17	8.81	0.36	1.81	10.05	0.25
6126- 8	R	1	9.08	0.4	1.34	8.84	0.24
6126- 39	R	0.97	14.89	0.36	1.46	9.06	0.23
6126- 16	R	1.07	8.23	0.34	1.2	8.71	0.28
6126- 51	R	0.97	14.89	0.36	1.5	10.93	0.31
6126- 62	R	1.15	8.33	0.39	1.32	10.46	0.35
6126- 67	R	1.08	9.2	0.24	1.59	9.56	0.38
6126- 118	R	1.04	8.68	0.32	1.28	8.69	0.24
Mean yellow	Y	1.05	10.93	0.37	1.4	9.42	0.24
Mean red	R	1.1	12.63	0.34	1.25	10.33	0.34
LSD 0.05	-	0.08	1.01	0.05	0.21	1.91	0.1
CDC Maxim	R	1.25	8.44	0.36	-	-	-
CDC Greenstar	Y	1.11	7.64	0.24	-	-	-
LSD 0.05		0.02	0.06	0.09	-	-	-

Vio- violaxanthin; Lut- lutein; Zea-zeaxanthin,Y- yellow, R- red

As in the previous two crosses, for cross 6126 lutein concentration was the highest among the three carotenoids for all genotypes in both the F_2 and F_3 generations. For the red × yellow cotyledon cross the genotypes 6126-110 and 6126-8 had the highest concentration of violaxanthin and zeaxanthin, respectively, and were similar to both the parents in the case of violaxanthin and to the red parent in for zeaxanthin. For lutein, the genotypes 6126-39 and 6126-51 had the highest concentration and were the most different from the other genotypes and the red parent in F_2 (Table 3.10).

In the F_3 generation, genotype 6126-5 had the highest violaxanthin concentration and was similar to both parents. For lutein, genotype 6126-27 had the highest concentration, similar to the red cotyledon parent. For zeaxanthin, 6126-67 had the highest concentration and was similar to red cotyledon parent like the F_2 (Table 3.10).

Table 3.11 Mean and variance components of carotenoid concentration for the three generation in cross 6126 between CDC Maxim (R) × CDC Greenstar (Y)

Cross	Trait	Generation	Mean±SD (µg g^{-1})	CV %	Range	F-values and Pr>F for genotype	Variance components	
							σ^2g	σ^2e
6126	Vio	Parents	1.18±0.10	6.83	1.11-1.25	3.01	0.01	3×10^{-3}
		F_2	1.08±0.11	3.26	0.97-1.32	18.70***	0.01	3×10^{-3}
		F_3	1.37±0.20	9.20	1.15-1.81	7.03***	19×10^{-3}	22×10^{-3}
	Lut	Parents	8.04±0.57	0.06	7.64-8.44	25328.50**	0.32	0.01
		F_2	12.16±6.86	3.73	8.23-14.89	456.83***	46.98	0.20
		F_3	9.75±1.28	11.39	8.69-12.88	5.02***	1.89	0.90
	Zea	Parents	0.30±0.08	2.48	0.24-0.36	234.98*	0.01	4×10^{-4}
		F_2	0.35±0.04	6.79	0.24-0.40	6.66***	2×10^{-3}	1×10^{-3}
		F_3	0.28±0.05	19.52	0.23-0.38	2.20	1×10^{-3}	3×10^{-3}

*, ** and *** Significant at the 0.05, 0.01 and 0.001 probability levels respectively; 6126- CDC Greenstar (Y) × CDC Maxim (R), Vio- violaxanthin, Lut- lutein, Zea- zeaxanthin.

In the red × yellow cotyledon cross 6126, no significant difference was detected for violaxanthin concentration between the two parents. In the F_2 and F_3 generations, significant differences in violaxanthin concentration among genotypes was observed (P<0.05). In each of the three generations significant differences in lutein concentration between genotypes was observed. For zeaxanthin concentration, the parent and F_2 generations had significant differences, while there was no significant difference between the genotypes of F_3 for this trait (P<0.05) (Table 3.1).

3.4 Discussion

Carotenoid concentration was phenotyped in populations of lentil that were specifically designed to segregate for three cotyledon colours. This is the first detailed study of the genetics of carotenoid concentration in lentil. The major carotenoids detected in lentil cotyledons observed in this study were lutein, violaxanthin and zeaxanthin. The mean total concentration of the three carotenoids was highest in green cotyledons compared to yellow and red cotyledons.

The significant variation in concentration of carotenoids at the segregation generations F_2 and F_3 confirms that there is segregation of genes governing the specific carotenoid concentrations.. Among the three carotenoids lutein concentration was highest compared to violaxanthin and zeaxanthin which had similar concentrations (about 10% of that of lutein).

In the biosynthesis of carotenoid pathway, lycopene is converted to α-carotene and β-carotene. α-carotene then converts to lutein, while β-carotene is converted to zeaxanthin, which in turn forms violaxanthin. The high concentration of lutein in all three lentil cotyledon types shows that there is a major flux in the biosynthetic pathway towards α-carotene branch which forms lutein.

Based on determination of the concentration of carotenoids for the specific cotyledon colours of lentil, it can be concluded that cotyledon colour is not a reliable predictor of carotenoid concentration level. Some red cotyledon genotypes in the F_3 generation from the red × green cotyledon cross had the same or higher total carotenoid concentration as that of the green cotyledon seeds in the yellow × green cotyledon cross. The specific cultivars involved in the cross had a greater effect on carotenoid concentration along with environmental effect.

According to the study on seed size and cotyledon colour, it was found that large seed size was associated with yellow cotyledon colour and small with red cotyledon colour (Fedoruk et al.,

2013). In the study conducted by Lazaro et al. (2001) it was found that of 101 Spanish landraces, yellow cotyledon was found to be present in 87% of the genotypes, while red-orange only occurred in 13% of the landraces. Roy et al. (2012) found that 90% of 110 lentil global accessions were red cotyledon and 10% were yellow cotyledon lentil accessions. Similar to Roy et al. (2012), Singh et al. (2014) found red cotyledon genotypes were more frequent that than yellow cotyledon types in a group of 405 global wild accessions of lentil germplasm. In terms of carotenoid concentration, red (10.05 μg g^{-1}) has higher carotenoids concentration than yellow cotyledon (8.99 μg g^{-1}). This could be because of a dilution effect on the carotenoid concentration in association with seed size.

In case of the green cotyledon lentils, which were medium to small size and had the highest carotenoid concentration, this could be explained as a combination of the environmental and genetic factors which induce smaller size (lower dilution effect) and higher expression of gene involved in carotenoid formation and storage. According to Kaliyaperumal et al. (2014) green cotyledon pea had higher carotenoid concentration than yellow cotyledon pea. A high correlation was found between lycopene content and colour value in study done in red tomato (Brandt et al., 2006).

It was observed that carotenoid concentration was higher in seeds produced on plants that experienced environmental stress compared to those produced on plants grown under favorable conditions (Demmig-Adam et al., 1992). This could be due to the plant differentiation in response to stress, thereby causing production of higher concentrations of pigments like chlorophyll and carotenoids for photosynthesis during seed development to tide over the unfavorable environmental conditions and the starch storage takes place towards the at seed maturity. Alternatively, an attenuation of ability to store starch during seed filling could result in a concentration effect.

Identification of the concentrations of the three different carotenoids in red green and yellow cotyledon produced in the three different generations in the three crosses under the same environment was the objective of the first study explained in chapter 3. Chapter 4 presents the results of the association analysis of carotenoid concentration with molecular marker. This work provides a framework for understanding the genetic basis of carotenoid concentration in lentil seeds.

CHAPTER 4

Study II: Association mapping of carotenoid colour and cotyledon colour in Lentil

4.1 Background

Lentil is a cool season pulse crop which has become an important part of the agricultural landscape in the northern prairies of North America. This high protein crop is widely consumed and has a growing global demand (Raghuvanshi, 2001). To meet the growing demand, production is steadily and rapidly increasing, especially in Canada.

One of the long term objectives in all breeding programs is to improve and maintain the nutritional quality of the product. Carotenoids are linked to human health for their pro-vitamin A and antioxidant properties. In the case of lentil, the potential for genetic improvement of carotenoid concentration has not been adequately studied.

Association mapping is considered a powerful high resolution tool for understanding the inheritance of complex quantitative traits like carotenoid concentration. The basic method is to associate trait with SNP markers on the basis of linkage disequilibrium (LD). A marker and a trait can be statistically associated if the marker is in LD with the gene responsible for the trait. By associating molecular markers with the phenotypic trait, quantitative traits can be mapped in lentil for the purpose of crop improvement. The use of molecular markers is limited in lentil breeding programs because until recently, many of the markers are not heritable in populations (Ford et al., 2009). Another reason is screening with markers can be both difficult and costly. A lentil association mapping (LAM) panel was formed to study and identify single nucleotide polymorphism markers (SNPs) that can be associated with phenotypic variations in the individual genotypes of the panel (Fedoruk, 2013).

4.1.1 Hypothesis and Objectives

By the analysis of the phenotypic data (violaxanthin, zeaxanthin and lutein concentration) specific carotenoids linked to SNPs (single nucleotide polymorphisms) can be identified for potential use in breeding lentil genotypes with higher concentration of carotenoids.

The objectives of this study were to identify the SNP markers associated with carotenoid concentration and cotyledon colour in lentil.

4.2 Materials and methods

4.2.1 Plant material

In this study 60 lentil accessions were grown at two locations, the SPG (Saskatchewan Pulse Growers) farm near Floral, SK and the Preston plot area in Saskatoon in 2011 and 2012. Genotypes were sown in 1 m^2 plots in randomized complete block design (RCBD) with three replicates in 2011 and 2012 at both locations SPG and Preston. Among the 60 genotypes, there were 8 breeding lines and 52 cultivars from around the world (Appendix Table 2). Fedoruk et al. (2013) conducted a study on the SNP marker loci associated with seed size and shape of lentil, and found that the most important QTL for seed diameter was located close to the locus *yc* that determines red/ yellow cotyledon colour. In their study, Fedoruk et al. (2013) were able to identify the genomic regions which are associated with controlling cotyledon colour at the *yc* locus.

4.2.2 Phenotyping of carotenoids for association mapping

The lentil seed samples were stored at room temperature. The cotyledon colour (red, yellow or green) for each sample was recorded. Samples of 0.1 g from three biological reps of 60 genotypes from each of the two locations and two years were ground to fine powder using an Udy mill (Cyclone Sample Mill, UDY Corporation, Fort Collins, Colorado, USA). Milled samples were placed in re-sealable plastic bags to prevent dampening. The plastic bags were placed inside brown envelopes to avoid breakdown of carotenoids by light. HPLC analysis was performed on each ground sample. The detailed protocol is described in Chapter 3. The solvent system was ACN (58%): MeOH (22%): DCM (20%) and the extraction buffer was ACN (50%): DCM (50%) + 0.01% BHT. About 0.1 g of the sample was placed in an Eppendorf tube, the extraction buffer was added and the tube was centrifuged for 15 min at 11,000 rpm. The HPLC peaks were recorded and the concentration of carotenoids for each peak in each sample was calculated. A C- 30 carotenoid column was used and the peaks were recorded and produced by diode array detector.

4.2.3 Phylogenetic Tree Construction

Nei's (1972) standard genetic distance measurement was used to calculate genetic distances using the software program SPaGeDi (Hardy and Vekemans, 2002). For accuracy, 1,000 individual bootstrap replications were performed. UPGMA (Unweighted Pair Group Method with Arithmetic Mean) was used to construct the phylogenetic tree and data were later visualized using the software Tree View (Page, 1996).

4.2.4 Population Structure and Kinship Calculations

The program STRUCTURE v2.2 (Pritchard, 2000) and admixture model, with a burn-in time of 50,000 were used to analyze the number of sub-populations in the panel. The model was set to a total of 100,000 Markov chain Monte Carlo repetitions and 1-8 number of K runs, each with five iterations. Based on the procedure illustrated by Evanno et al. (2005) the number of groups was decided. After submitting each of the values of K, the group with the highest ad hoc statistic K value (Earl et al., 2012) as returned by the STRUCTURE harvester website (Pritchard et al., 2000) was selected.

4.2.5 Association Analysis

For association mapping (AM) analysis, the TASSEL version 3.0 (Bradbury et al. 2007) program was used. A generalized linear model (GLM) was used to find Q which was the only covariate set. The significant level was customized according to the false discovery rate (FDR) correction. Values below <0.05 *P*-value was considered significant.

To identify the candidate gene for the markers, the unique name for each marker was run in the Knowpulse portal (http://knowpulse.usask.ca/portal/search/markers) from which the contig sequence was identified in *Medicago truncatula*. The identified specific sequence was blasted to identify the function that the sequences have in the different species in order to help identify possible associated candidate genes (http://plantgrn.noble.org/LegumeIP/search.do).

4.2.6 Statistical Analysis

Statistical analysis was performed using SAS 9.3 (SAS Institute Inc. 2011. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.). The program was run for location (SPG and Preston), and years (2011 and 2012). Bartlett's homogeneity test was used to determine if data

from the two locations from two years could be combined. Then, data were combined over years and locations, and finally, all years and locations were combined where both were random factors and their interaction was considered as random factor as well. PROC GLM (generalized linear model) was used to analyze the phenotypic data because of the design of the experiment.

4.3 Results

4.3.1 Phenotypic Data

From the ANOVA for the effect of genotype it was concluded that there were significant difference among the genotypes (Table 4.1) for the three carotenoid traits – zeaxanthin, lutein and violaxanthin concentration in lentil seeds. The ANOVA for the effect of location by genotype and year by genotype interaction was significant for all traits. The ANOVA for the three-way interaction among year by location by genotype was significant only for violaxanthin. Due to the small population size, the interaction between the three carotenoids and the factors genotypes, year and location for the two years and two locations were analyzed together.

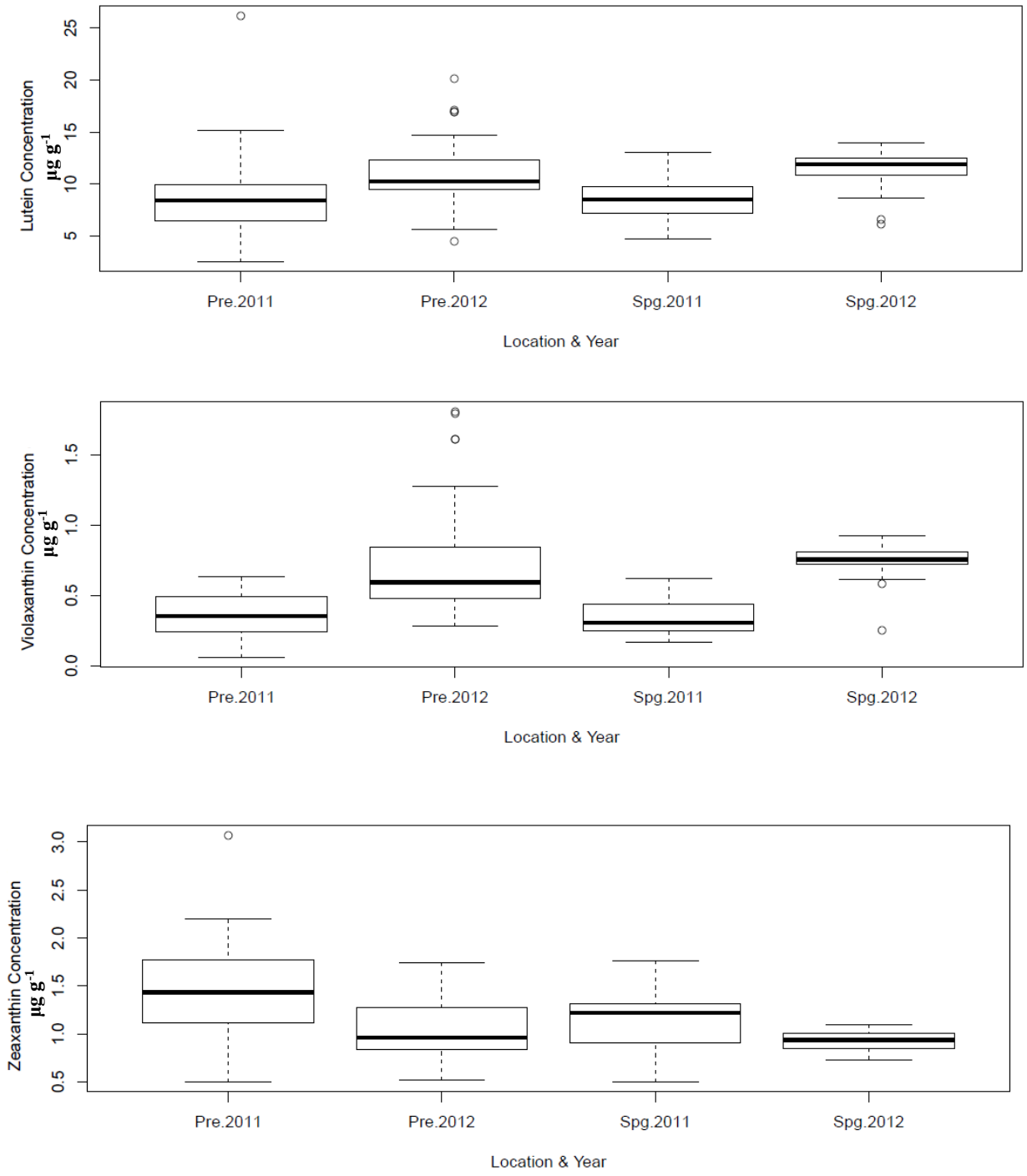
Table 4.1 F-values for the ANOVA of zeaxanthin, lutein and violaxanthin concentration in lentil seeds for 60 lentil accessions in the association mapping panel

		F-value		
Effect	df	Zeaxanthin	Lutein	Violaxanthin
Genotype	59	8.11***	12.3***	6.74***
Loc*Genotype	59	3.92***	8.01***	6.48***
Year*Genotype	59	8.45***	7.00***	7.05***
Year*Loc*Genotype	59	2.32	2.51	4.84***

*** P≤ 0.001, df- degree of freedom

In the Figure 4.1 the three box plots represent the variation in concentration among the accession at different locations and years. For both lutein and violaxanthin, locations Preston 2011 and 2012 showed higher variability in concentration among genotypes than at location SPG for both years. For zeaxanthin we can see that genotypes at the location Preston in 2011 had the highest variation in concentration and the least in SPG in 2012.

Figure 4.1 Concentration of the carotenoids violaxanthin, lutein and zeaxanthin in samples of lentil seeds grown in the lentil association mapping panel for 2011 and 2012 at Preston and SPG locations



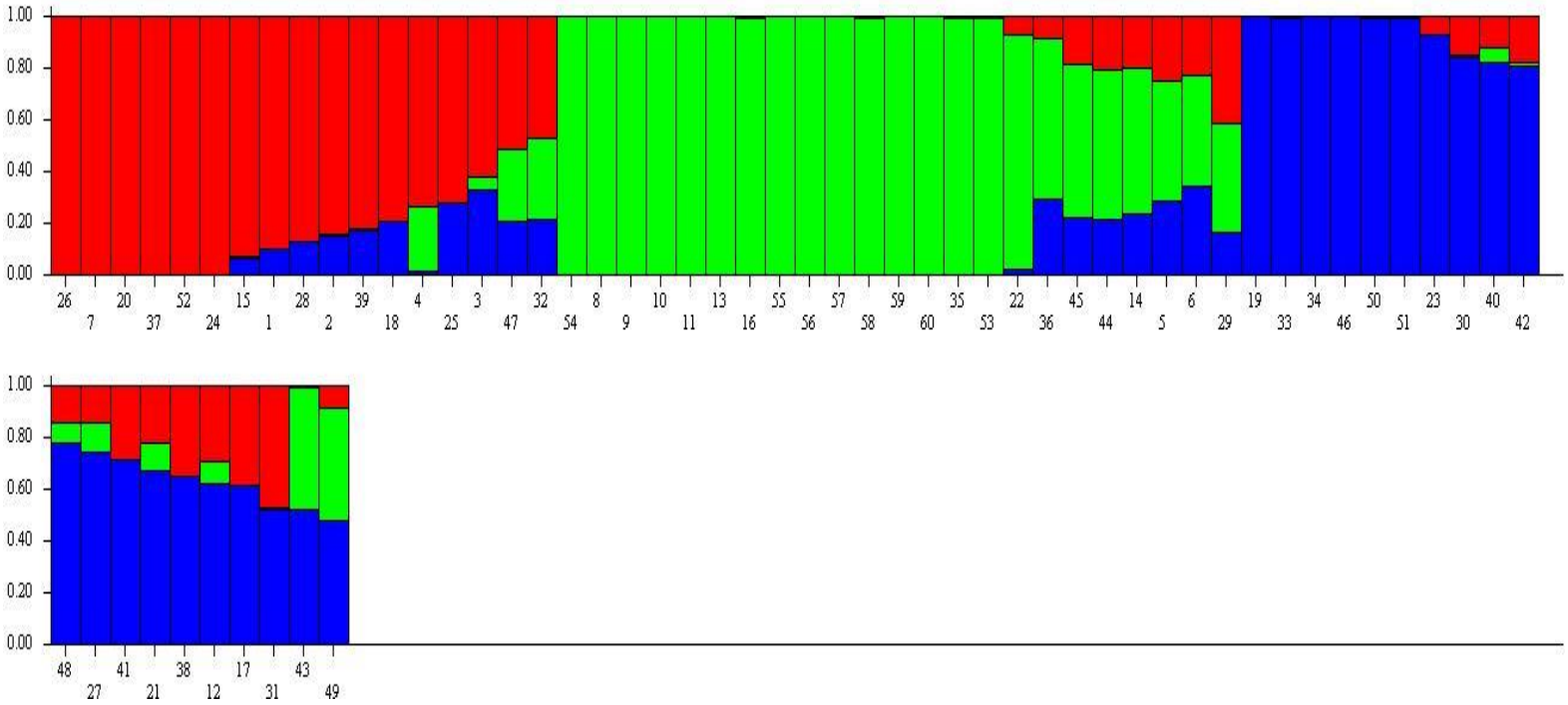
Pre-Preston, Spg-SPG (Saskatchewan Pulse Growers)

4.3.2 Population Structure

The population structure of the LAM panel samples was analyzed using the STRUCTURE (Web v0.6.94), program which determined the number of clusters in the LAM samples at the highest ad hoc K value K= 3 (Figure 4.2). The detailed list of individuals in each cluster group and the level of admixtures among them are listed in Appendix 2. Group 2, the largest (shown in green) contained 38.3% of the lines, while Group numbers 1 (red) and 3 (blue) had 28.3% and 33.3% of the genotypes, respectively. The genotypes in all three groups were both breeding lines and cultivars. The three cluster groups were separated based on carotenoid concentration in the seeds and on the breeding history of the genotypes. The mean total carotenoid concentration for Group 1 was 11.06 (µg g^{-1}), 10.8 (µg g^{-1}) for Group 2 and 12.31 (µg g^{-1}) for Group 3.

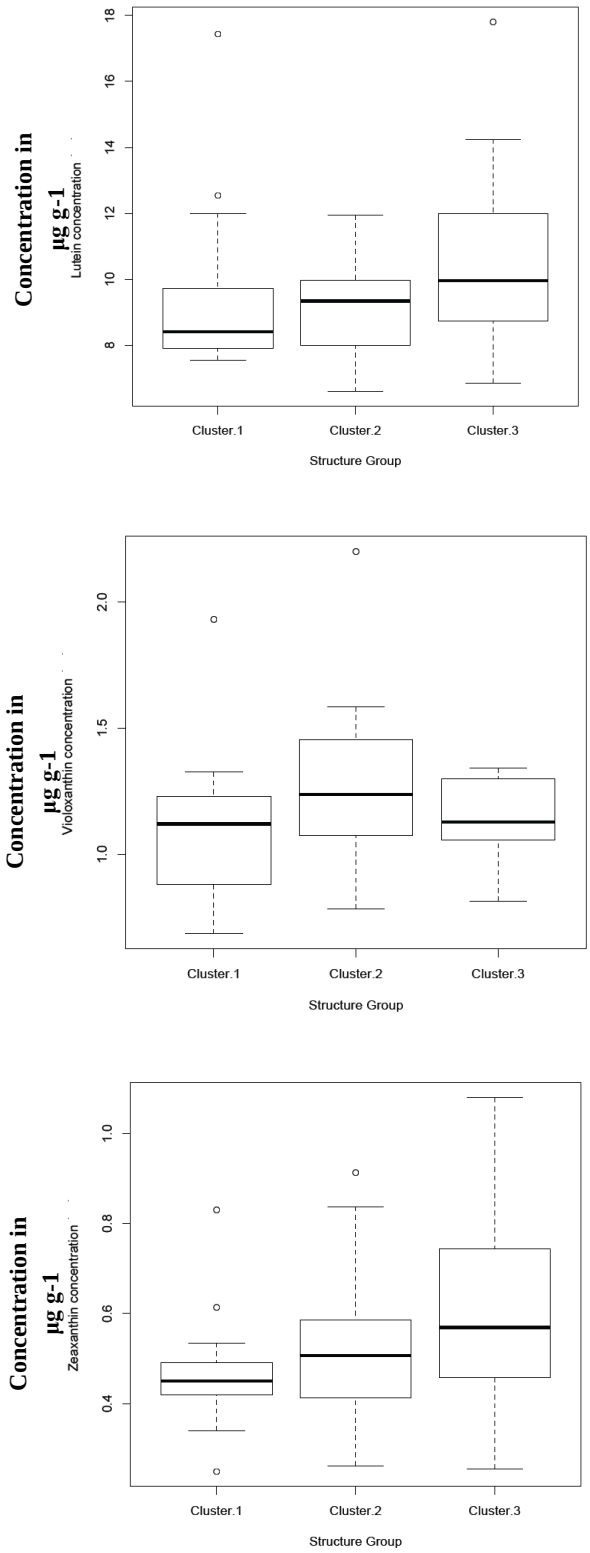
Each vertical bar in Figure 4.2 represents the individual while the colour represents the subpopulation and admixture of each of the 60 lentil lines genotyped and sorted into populations based on STRUCTURE analysis. Red represents cluster 1, green cluster 2 and blue cluster 3 and each numbered bar represents a genotype.

Figure 4.2 60 Lentil genotypes sorted into 3 population groups based on STRUCTURE analysis



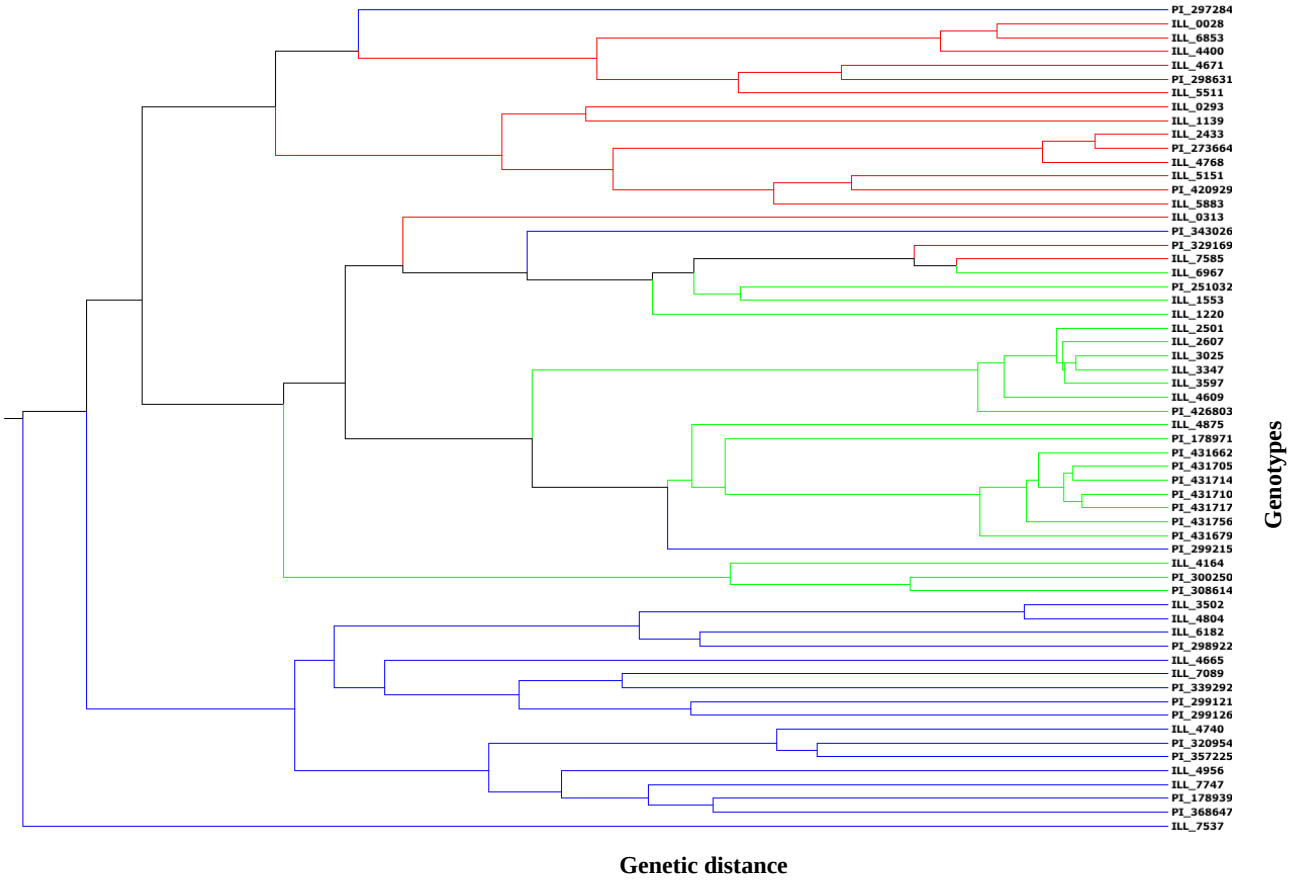
Numbers represent genotypes

Figure 4.3 Carotenoid concentrations in seeds of lentil for three genotypic clusters identified by STRUCTURE analysis



These three box plots represent the variation in the three carotenoid concentrations for the three cluster groups. For lutein and zeaxanthin cluster 3 genotypes had greater variability than cluster 1 and cluster 2. For violaxanthin, cluster 2 showed greater variation than clusters 1 and 3.

Figure 4.4 Dendogram generated by the UPGMA cluster analysis for the population sub-groups of 60 genotypes



The UPGMA clustering analysis in Figure 4.4, which is based on genetic distance, also indicated that the lines were grouped mainly by their carotenoid concentration. Each colour represents a population sub-group: red for sub-group 1, green for sub-group 2 and blue for sub-group 3. Some o genotypes have a different colour and belong to a different sub-group but appear between genotypes of another sub-group. This occurs because the colour of the genotype indicates the sub-group to which they belong while appearance in another colour sub-group

means that the genotype has some influence from this genepool of sub-groups. For example, genotype PI 297284 belongs to sub-group 3 (blue) but in the dendogram it appears with sub-group 1 (red). In the structure, this genotype is an admixture of cluster 1 (red, 0.34) and cluster 3 (blue, 0.65).

4.3.5 Association Analysis

For association mapping by TASSEL (Bradbury et al. 2007), the generalized likelihood method (GLM) was used. In the GLM model, the population structure of the LAM panel sample was taken into consideration. According to the study by Fedoruk (2013) seed diameter, seed thickness and seed plumpness, 31 different associations were determined at two different locations. All locations and years were combined for the analysis. In our study, the maximum likelihood method (MLM) analysis resulted in no significant association for any of the locations and years. Appendix 3 shows the significant markers in decreasing order of *P*-value meaning association with the carotenoids.

4.4 Discussion

In this study AM was used to identify SNP markers associated with carotenoid concentration and cotyledon colour in a panel with genotypes that are cultivars, breeding lines or landraces. Three sub-groups were observed in the panel using population structure (Figure 4.4). When the cotyledon colours of each of the accession were compared in each of the three groups it was found that not all genotypes in each group had the same cotyledon colour. For example Group 1 had nine red and eight yellow cotyledon accessions. In Group 2, there were 16 red, six yellow and one green cotyledon accessions. In Group 3 there were seven red, 12 yellow and one green cotyledon accessions. Here in the cluster 1, the number of red cotyledon accessions was slightly higher than the number of yellow types. In Group 2, almost 75% of the accessions had red cotyledons. In Group 3, in contrast to the first two groups, the number of yellow cotyledon accessions was higher than for red cotyledons. One green cotyledon accession was present in both Groups 2 and 3.

The division of the 60 accessions into three sub-groups was supported by the UPGMA tree analysis. The clustering pattern was similar to that of the sub-groups revealed by the population structure analysis. Some lines in the dendrogram did not group in the cluster as they did in the

structure sub groups. In the study by Liu et al. (2008) using 440 accessions of lentil (Chinese National Gene Bank), eight groups were found, and these were separated based mainly on geographic origin. Similarly, geographic origin of the accessions was found to be the major cause for variability among lentil germplasm in a study by Alabboud et al. (2009). The influence of geographic origin can also be observed in our study using the genetic diversity in the population structure groups. In the first structure group most of the accessions were from the Middle-East, while most accessions in group two were from South Asian countries and the accessions in the third group are mostly from the European countries. The study conducted by Bakhsh et al. (2013) confirmed again that traits like cotyledon colour, among other traits like testa pattern and anthocyanin pigmentation on seedlings to be monogenic in inheritance, and that red cotyledon was dominant to yellow cotyledon.

Both the GLM and MLM analyses were run after the markers were corrected to reduce false associations. Markers were found to be significantly associated to carotenoid concentration by GLM but none by the MLM method. This result was similar to that of the association mapping study conducted by Fedoruk (2013). MLM is a more conservative and stringent model than GLM because it also includes a correction based on kinship, while GLM does not. Comparing the two models, Neumann et al. (2011) found that the models were trait dependent suggesting that neither model is a perfect fit for every trait and species.

Using the GLM model, 168 markers were found to be significant after the *P*-value was corrected. The most significant marker, LcC03973p328, was associated with lutein concentration, but it did not map in the lentil LR-18 population map (Sari, 2014) which is a recombinant inbred line, from cross between CDC Robin and 964a- 46 (Appendix 3). A homolog of this marker was located by using BLAST with *Medicago truncatula* (Medtr8g042040.1) after comparing the contig sequence. The marker LcC03973p328, significant only for lutein, mapped 1653 Kbp away from the *Arabidopsis LUT1* gene homolog in *Medicago truncatula*. This gene encodes cytochrome P450-type mono- oxygenase which is part of a family of enzymes responsible for forming the $\mathcal{E}$-ring hydroxylation in the carotenoid biosynthetic pathway that catalyzes the formation of lutein from α- carotene (Tian et al., 2004). No contig sequence was mapped near the *LUT1* gene for significance association for violaxanthin. There is a potential for this marker to be used to select

for lutein and coincidently to select for zeaxanthin. This marker showed significant association for both lutein and zeaxanthin but none for violaxanthin.

Most of the markers were not polymorphic. This could be advantageous since the loci may not be significantly associated with the trait in the linkage map in AM. A total of 27 markers for lutein, 91 for violaxanthin and 50 for zeaxanthin out of 1388 markers showed significant association, of these 6 markers are associated with all three traits (Appendix 3).

CHAPTER 5

General Discussion

Conclusions and Future Work

From the results it could be observed that green cotyledon lentils have higher carotenoid concentration and that SNP markers could be associated with concentration of specific carotenoids using association mapping. This would result in the ability to design a lentil genetic improvement program so that both cotyledon colour and markers could be used to select for increasing specific carotenoid concentrations in lentil germplasm and cultivars.

According to Delgado-Vargas (2010) lutein is the most abundant carotenoid in nature. In the first study lutein was found to be the most abundant carotenoid in lentil cotyledons and zeaxanthin and violaxanthin were present at much lower concentrations (less than 10% of the total concentration). This was also observed in a previous study of chickpea (Abbo et al., 2005). Mean total carotenoid concentration was approximately 27% higher in green cotyledon lentils compared to red cotyledon lentils, which in turn had higher concentration than yellow cotyledon lentils (red higher by ~8% compared to yellow). In a recent study, a similar result was observed for green cotyledon pea seeds which had mean carotenoid concentration higher compared to yellow pea seeds (Kaliyaperumal et al., 2014). In the case of fruit tissues, for example in bell peppers, it was observed that red bell peppers had more carotenoids that green (Zhang et al., 2003).

Since the difference in the carotenoid concentration is quantitative rather than qualitative for the individual parents in the three crosses, the variation in the concentration of carotenoids can be explained as the variation caused by expression of genes that regulate and determine the rate of carotenoid formation and degradation. The expression of the ‘stay-green’ gene was studied in model legume plant *Medicago truncatula*. It causes retention of about 50% of chlorophyll at senescence in leaves, pods and cotyledon. This can be explained as the result of high concentration of carotenoids at maturity, which protects chlorophyll from sun damage and breakdown (Zhou et al., 2011). This could also be the reason that green cotyledon lentil plants maintain green cotyledons, green pods and green foliage at seed maturity.

The difference in the concentration of zeaxanthin and violaxanthin in the same year and location can be explained by understanding the biosynthetic pathway of carotenoids (Figure 2.1) which shows that the breakdown of zeaxanthin forms violaxanthin. For the traits lutein and zeaxanthin the relationship between the two in terms of concentration showed transgressive segregation (Abbo et al., 2005).

Both MLM and GLM models were performed in this study, but none of the markers were significantly associated for the three traits of interest using MLM while for GLM 168 markers found to be linked to the traits. The MLM results coincided with the results of the Fedoruk (2013) study as the 60 genotypes used in study are a sub- set from his study. Of the 168 markers that were found to be significantly associated, maker LcC03973p328 was found to be associated with both lutein and zeaxanthin. Each SNP marker is developed from a contig sequence candidate genes of traits using the BLAST program to identify the homologous sequence in the model crop *Medicago truncatula.* These were mapped near the candidate gene *Lut1* which code for cytochrome P450 family of enzymes which are essential in the synthesis of lutein in the biosynthetic pathway. Among the three clusters, lentil genotypes in cluster 3 showed high concentration and variability for lutein and zeaxanthin while for zeaxanthin cluster 2 showed higher concentrations and variability. When the variability and concentration was compared between the genotypes in the first study and LAM panel samples, genotypes from the first study involving the crossing between different cotyledon colour lentil lines and the progenies were found to have higher variability and concentration.

Many approaches have been made in order to increase carotenoid concentration which can also be implemented in lentils like suppressing / blocking the branching from lycopene in the biosynthetic pathway towards lutein and promoting the production of β carotene by the over expression of the *β cyclase* enzyme this mutation can be brought about by employing radiation mutation techniques, or as done in rice (daffodil- Golden rice- 1, maize- Golden rice- 2) replacing the *phytoene synthase* with the maize or daffodil enzyme as it is the rate limiting step in the pathway along with enzymes *phytoene desaturase* and *β cyclase* from source with high carotenoid concentration and compatible genome (Beyer et al., 2002).

Further selection and breeding could increase the concentration of dietary carotenoids in lentils, bringing it closer to the recommended daily carotenoid consumption that is about 5-6 mg of

carotenoids per day for an adult (depending on age, gender and body mass index) which equals about 5-6 servings of vegetables and fruits (NCI, USA). It would be informative to study the carotenoid concentration in the progenies from the cross between these high and low concentration lentil genotypes. Despite the variation observed in carotenoid concentration, lentil cotyledons had higher carotenoid concentration than that found in dry "golden rice" endosperm (1.6 µg g^{-1}; Ye et al., 2000). In order to increase the bioavailability, according to Chitchumroonchokchai et al. (2009) the unsaturated fatty acids could be increased as it is shown to be the promoter of carotenoids in the experiments done *in vitro*.

According to Abbo et al. (2005), chickpea seeds had higher carotenoid concentration than golden rice endosperm. In countries where fruits and vegetables are scarce or only seasonally available and pulses like lentil and chickpea are a prominent dietary staple, these pulses could be a superior source of dietary carotenoids compared to cereals. When we calculate the carotenoid content of 50 g of green cotyledon lentil (CDC QG- 2) consumed by an adult, which has 12.80 µg g^{-1} carotenoid concentration, that is 640 µg 50 g^{-1} of carotenoid concentration. This amount may not be able to solve the vitamin A deficiency yet, but it is a potential source to help resolve it in the long run. Even though these initiatives in carotenoid enrichments in staple crops are not able to solve the entire problem of vitamin A deficiency, it is still a great leap for mankind since it could help in reducing the proportion of the population that suffers from disorders associated with vitamin A deficiency like nyctalopia, xerophthalmia etc. As long as there is enough vitamin A concentration in the human body to avoid these health problems, it could be sufficient even if intake is a little lower than the RDA.

It would be beneficial to screen more lentil accessions for specific carotenoid concentration so that the novel sources of high carotenoid alleles can be identified and used for breeding to broaden the present gene pool. Molecular breeding would also be another step forward in developing lentil lines with high carotenoids concentration. Validation of the linked marker to carotenoids in different genetic backgrounds needs to be done in future.

References

1. Abbo S., Molina C., Jungmann R., Grusak M.A., Berkovitch Z., Reifen R., Kahl G., Winter P., Reifen R., 2005. Quantitative trait loci governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). Theoretical and Applied Genetics 111: 185–95.
2. Alabboud I., Szilagye L., Roman G.V., 2009. Assessment of genetic diversity in lentil (*Lens culinaris* Medik.) as revealed by RAPD markers. Scientific Papers Series A 2: 439–44.
3. Aljohi H., Dopler-Nelson M., Wilson T., 2014. Consumption of 12 eggs per week for 1 year increases serum zeaxanthin concentrations but not other major carotenoids, tocopherols, and retinol in humans. The FASEB Journal 28: 645.25.
4. Armstrong G., Hearst J.E., 1996. Carotenoids: Genetics and molecular biology of carotenoid pigment biosynthesis. The FASEB journal 10: 228–37.
5. Bakhsh A., Iqbal M., Cheema N.M., 2013. Inheritance of morphological characters associated with plant and dried seeds in lentil (*Lens culinaris* M.). Pakistan Journal of Botany 45: 1497– 1502.
6. Beyer P., Al-Babili S., Ye X., Lucca P., Schaub P., Welsch R., Potrykus I., 2002. Golden Rice: Introducing the β-carotene biosynthesis pathway into rice endosperm by genetic engineering to defeat vitamin A deficiency. The Journal of Nutrition 132: 5065–5105.
7. Bhatty R. S., 1989. Cooking quality and losses of phytic acid, calcium, magnesium and potassium of lentils soaked in different solutions. Canadian Institute of Food Science and Technology Journal 22: 450– 455.
8. Botella-Pavía P., Besumbes O., Phillips M.A., Carretero-Paulet L., Boronat A., Rodríguez-Concepción M., 2004. Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in Controlling the supply of plastidial isoprenoid precursors. The Plant Journal 40: 188–99.
9. Bradbury P.J., Zhang Z., Kroon D.E., Casstevens T.M., Ramdoss Y., Buckler E.S., 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–35.

10. Brandt S., Pék Z., Barna E., Lugasi A., Helyes L., 2006. Lycopene content and colour of ripening tomatoes as affected by environmental conditions. Journal of the Science of Food and Agriculture 86: 568– 72.
11. Breseghello, F., Sorrells M.E., 2006. Association analysis as a strategy for improvement of quantitative traits in plants. Crop Science 46: 1323– 30.
12. Britton G., 1995. Structure and properties of carotenoids in relation to function. The FASEB Journal 9: 1551– 58.
13. Cadenas E., Packer L., 2002. Handbook of Antioxidants. 2nd. Marcel Dekker, Inc., New York.
14. Chaudhary N., Nijhawan A., Khurana J.P., and Paramjit Khurana P., 2010. Carotenoid biosynthesis genes in rice: Structural analysis, genome-wide Expression profiling and phylogenetic analysis. Molecular Genetics and Genomics 283: 13– 33.
15. Chitchumroonchokchai C., Kamonpatana K., Ferruzzi M.G., Harrison E.H. Failla M.L., 2009. Unsaturated fatty acids promote carotenoid bioavailability in vitro. The FASEB Journal 896: 9.
16. Clinton S.K, Emenhiser C., Schwartz S.J., Bostwich D.G., Williams A.W., Moore B.J., Erdman J.W., 1996. Cis-trans lycopene isomers, carotenoids, and retinol in the human prostate. Cancer Epidemiology, Biomarker and Prevention 5: 823– 33.
17. Clotault J., Peltier D., Berruyer R., Thomas M., Briard M., Geoffriau E., 2008. Expression of carotenoid biosynthesis genes during carrot root development. Journal of Experimental Botany 59: 3563– 73.
18. Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B., Pang E.C.K., 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142: 169– 96.
19. Cunningham F.X ., Pogson B., Sun Z., McDonald K.A., Penna D.D., Gantt E., 1996. Functional analysis of the beta and epsilon lycopene cyclase enzymes of Arabidopsis reveals a mechanism for control of cyclic carotenoid formation. PlantCell 8: 1613– 26.
20. Cunningham, F. X., E. Gantt E., 1998. Genes and enzymes of carotenoid biosynthesis in plants. The Plant Physiology and Plant Molecular Biology 49: 557– 83.

21. Demmig-Adams B., AdamsW.W., 1992. Photoprotection and other responses of plants to high light stress.Annual Review of Plant Physiology and Plant Molecular Biology 43: 599– 26.
22. Dong H., Deng Y., Mu J., Lu Q., Wang Y., XuY., Chu C., Chong K., Lu C., Zuo J., 2007. The Arabidopsis spontaneous *Cell Death*1 gene, encoding a Zeta-carotene desaturase essential for carotenoid biosynthesis, is involved in chloroplast development, photoprotection and retrograde signalling. Cell Research 17: 458– 70.
23. Dutta D., Chaudhuri U.R., Chakraborty R., 2005. Structure, health benefits, antioxidant property and processing and storage of carotenoids. African Journal of Biotechnology 4: 1510– 20.
24. Dwamena C.A., McGhie T., Wibisono R., Montefiori M., Hellens R.P., Allan A.C., 2009. The kiwifruit lycopene beta-cyclase plays a significant role in carotenoid accumulation in fruit. Journal of Experimental Botany 60: 3765– 79.
25. Earl D.A., 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetic Resources 4: 359– 61.
26. Eliassen A.H, Liao X., Tworoger S.S., Hankinson S.E., 2014. Plasma carotenoids and risk of breast cancer over 20 years of follow-up in the nurses' health study. Proceedings of the 105th Annual Meeting of the American Association for Cancer Research, San Diego, 19: – 2919.
27. Ellison S.L., Iorizzo M., Senalik D.A., Simon P.W., 2014. Orange is the new yellow: cracking the genetic code controlling carotenoid accumulation in carrot (*Daucus carota* L.). Plant and Animal Genome XXII: 1212: 680.
28. El-Qudah J.M., 2014. Estimation of carotenoid content of selected Mediterranean legumes by HPLC. World Journal of Medical Sciences 10: 89– 93.
29. Emami M. K., Sharma B., 1996. Confirmation of digenic inheritance of cotyledon colour in lentil (*Lens culinaris*). The Indian Journal of Genetics and Plant Breeding 56: 536– 68.
30. Evanno G., Regnaut S., Goudet J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14: 2611– 20.
31. FAOSTAT, 2014. http://faostat.fao.org/. Accessed December, 2014.

32. Fedoruk M.J., Vandenberg A., Bett K.E., 2013. Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. The Plant Genome 6: 3.

33. Fedoruk M.J., 2013. Linkage and association mapping of seed size and shape in lentil. Master thesis. University of Saskatchewan, Canada.

34. Fernandez-Marin B., Milla R., Martin-Robles N., Arc E., Kranner I., Becerril J.M., Garcia-Plazaola J.I., 2014. Side-effect of domestication: Cultivated legume seeds contain similar tocopherols and fatty acids but less carotenoids than the wild counterpart. BMC Plant Biology 14: 1599.

35. Ford R., Mustafa B., Baum M., Rajesh P.N., 2009. Advances in molecular research. In: Erskine W., Muehlbauer F.J., Sarker A., Sharma B. (eds), The lentil: botany, production and uses. Cameroon Journal of Experimental Biology, pp155– 71.

36. Fulton T.M., Grandillo S., Beck-Bunn T., Friedman E.J., Frampton A., Lopez J., Petiard V.,Uhli J., Zamir D., Tanksley S.D., 2000. Advanced backcross QTL analysis of a lycopersicon esculentum X lycopersicon parviflorum cross. Theoretical and Applied Genetics 100: 1025– 42.

37. Gil R., Cohen M., Zamir D., and Hirschberg J., 1999. Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant *Delta*. The Plant Journal 17: 341– 51.

38. Giovannucci E., Clinton S.K., 1998. Tomatoes, lycopene, and prostate cancer prostate. Proceedings of the Society for Experimental Biology and Medicine 218: 129– 39.

39. Hardy O.J., Vekemans X., 2002. SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618– 20.

40. Hata T.R., Scholz T.A., Ermakov I.V., McClane R.W., Khachik F., Gellermann W., Pershing L.K., 2000. Non-invasive raman spectroscopic detection of carotenoids in human skin. Journal of Investigative Dermatology 115: 441– 48.

41. Jiang G., 2013. Molecular markers and marker-assisted breeding in plants. In: Andersen S.B.(eds), Plant Breeding from Laboratories to Fields, pp.45– 83.

42. Kato M., Ikoma Y., Matsumoto H., Sugiura M., Hyodo H., Yano M., 2004. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. Plant Physiology 134: 824–37.
43. Kaliyaperumal A.K., Tar'an B., Diapari M., Arganosa G., Warkentin T.D., 2014. Effect of cultivar and environment on carotenoid profile of pea and chickpea. Crop Science 54: 2225– 35.
44. Kearsey M. J., Farquhar A.G.L., 1998. QTL analysis in plants: Where are we now?. Heredity 80: 137– 42.
45. Kimura M., Kobori C.N., Rodriguez-Amaya D.B., Nestel P., 2007. Screening and HPLC methods for carotenoids in sweetpotato, cassava and maize for plant breeding trials. Food Chemistry 100: 1734– 46.
46. Krinsky N.I.,1994. The biological properties of carotenoids. Pure and Applied Chemistry 66: 1003– 10.
47. Krinsky N.I., Johnson E.J., 2005. Carotenoid actions and their relation to health and disease. Molecular Aspects of Medicine 26: 459– 516.
48. Kumar P., Gupta V.K., Misra A.K., Modi D.R., Pandey B.K., 2009. Potential of Molecular Markers in Plant Biotechnology. Plant Omics 2: 141– 62.
49. Lachman J., Hejtmánková K., Kotíková Z., 2013. Tocols and carotenoids of einkorn, emmer and spring wheat varieties: Selection for breeding and production. Journal of Cereal Science 57: 207– 14.
50. Landrum J.T., Bone R.A., Joa H., Kilburn M.D., Moore L.L., Sprague K.E., 1997. A one year study of the macular pigment: The effect of 140 days of a lutein supplement. Experimental Eye Research 65: 57– 62.
51. Lázaro A., Ruiz M., de la Rosa L., Martín I., 2001. Relationships between agro/morphological characters and climatic parameters in Spanish landraces of lentil (*Lens culinaris* Medik.). Genetic Resources and Crop Evolution 48: 239– 49.
52. Liu J., Guan J.P., Xu D.X., Zhang X.Y., Gu J., Zong X.X., 2008. Genetic diversity and population structure in lentil (*Lens culinaris* Medik.) germplasm detected by SSR markers. Acta Agronomica Sinica 34: 1901– 09.
53. Miles C.M., Wayne M., 2008. Quantitative trait locus (QTL) analysis. Nature Education 1: 1– 7.

54. Mishra S.K., Sharma B., Sharma S.K., 2007. Genetics and Cytogenetics of Lentil. In: Yadav S.S., McNeil D.L., Stevenson P.C., (eds.), Lentil, pp.187– 208.

55. Mortensen L.H.,1999.Endothelin and the central and peripheral nervous systems: A decade of endothelin research. Clinical and Experimental Pharmacology and Physiology 26: 980– 84.

56. Muller H., 1996. Daily intake of carotenoids (carotenes and xanthophylls) from total diet and the carotenoid content of selected vegetables and fruit. Zeitschrift für Ernährungswissenschaft 35: 45– 50.

57. Nelles G., Dias I., Weber D., Li L., Stahl W., Grune T., Polidori M.,Griffiths H., 2014. Plasma levels of HDL and carotenoids are lower in alzheimer's disease patients with vascular comorbidities but plasma protein nitration and oxidation is unaffected. American Academy of Neurology 82: 205.

58. Nei M., 1972. Genetic distance between populations. American Naturalist 106: 283– 92.

59. Neumann K., Kobiljski B., Denčić S., Varshney R K., Börner A., 2011. Genome-wide association mapping: A case study in bread wheat (*Triticum aestivum* L.). Molecular Breeding 27: 37– 58.

60. Oliveira D., Alves G., Castilhos F.D., Renard C.M.G.C., Bureau S., 2014. Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. Food Research International 60: 154– 62.

61. Ooijen J.W.V., Kyazma B.V., 2006. JoinMap® 4. Software for the calculation of genetic linkage mapin experimental populations, Wageningen.

62. Oplinger E.S., Hardman L.L., Kaminski A.R., Kelling K.A., Doll J.D., 1990. Lentil. Alternative Field Crops Manual. University of Wisconsin.

63. Oraguzie N.C., Wilcox P.L., Rikkerink E., 2007. Linkage disequilibrium. In: Oraguzie N.A., Rikkerink E.H.A., Gardiner S.E., Silva N. Association mapping in plants. Springer: 10- 39.

64. Owens B.F., Lipka A.E., Lundback M.M.,Tiede T., Diepenbrock C.H., Kandianis C.B., Kim E., Cepela J., Hernandez M.M., Buell C.R., Buckler E.R., Penna D.D., Gore M.A., Rocheford T., 2014. A foundation for provitamin A biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. Genetics 192: 1699–- 1716.

65. Page R.D., 1996. TreeView. An application to display phylogenetic trees on personal computer. Applied Biology Science 12: 357– 358.
66. Paiva S.A.R., Russell R.M., 1999. Beta-carotene and other carotenoids as antioxidants. Journal of the American College of Nutrition 18: 426– 33.
67. Potrykus I., 2001. Golden rice and beyond. Plant Physiology 125: 1157– 1161.
68. Pritchard J.K., Stephens M., Donnelly P., 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945– 59.
69. Qudah E., Muhammad J., 2009. Identification and quantification of major carotenoids in some vegetables. American Journal of Applied Sciences 6: 492– 97.
70. Raghuvanshi R.S., Singh R., Singh R., 2001. Nutritional composition of uncommon foods and their role in meeting micronutrient needs. International Journal of Food Sciences and Nutrition 52: 331– 35.
71. Ramachandran A., Pozniak C.J., Clarke J.M., Singh A.K., 2010. Carotenoid accumulation during grain development in durum wheat. Journal of Cereal Science 52: 30– 38.
72. Rao A.V., Agarwal S., 1999. Role of lycopene as antioxidant carotenoid in the prevention of chronic diseases: A review. Nutrition Research 19: 305– 23.
73. Ray H., Bett K., Tar'an B., Vandenberg A., Thavarajah D., Warkentin T., 2013. Mineral micronutrient content of cultivars of field pea, chickpea, common bean, and lentil grown in Saskatchewan, Canada. Crop Science 54: 1698– 1708.
74. R Development Core Team, 2011. R: A language and environment for statistical computing. R foundation for statistical computing. http://R-project.org/. Accessed November, 2011.
75. Ribaut J. M., Hoisington D., 1998. Marker-assisted selection: New tools and strategies. Trends in Plant Science 3: 236– 39.
76. Ronen G., Cohen M., Zamir D., Hirschberg J., 1999. Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. The Plant Journal 17: 341– 351.

77. Rouseff R., Raley L., 1996. Application of Diode Array Detection with a C-30 Reversed phase column for the separation and identification of saponified orange juice carotenoids. Journal of Agriculture and Food Chemistry 44: 2176– 81.
78. Rice-Evans C., Sampson J., Bramley P.M., Holloway D.E., 1997. Why do we expect carotenoids to be antioxidants *in vivo*? Free Radical Research 26: 381– 98.
79. Rodríguez G.R., Moyseenko J.B., Robbins M.D., Morejón N.H., Francis D.M., Knaap E.V., 2010. Tomato analyzer: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. Journal of Visualized Experiments 37: 1– 9.
80. Roy S., Islam M.A., Sarker A., Ismail M.R., Rafii M.Y., Mondal M.M.A., Malek M.A., 2012. Morphological characterization of lentil accessions: Qualitative characters. Bangladesh Journal of Botany 41: 2.
81. Sajjad M., Khan S.H., Ahmad Q., Rasheed A., Mujeeb-kazia., Khan I.A., 2012. Association mapping identifies QTLS on wheat chromosome 3A for yield related traits. Cereal Research Communications 42: 177– 88.
82. Santos C., Simon P., 2002. QTL analyses reveal clustered loci for accumulation of major provitamin A carotenes and lycopene in carrot roots. Molecular Genetics and Genomics 268: 122– 29.
83. Sari E., 2014. Genetics of resistance to ascochyta blight in lentil. PhD thesis. University of Saskatchewan, Canada.
84. Sauvage C., Segura V., Bauchet G., Stevens R., Do P.T., Nikoloski Z., Fernie A.R., Causse M., 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiology 165: 1120– 32.
85. SAS Institute Inc., 1988. SAS/Af User's Guide, Release 6. SAS.
86. Saskatchewan Ministry of Agriculture (SMA). Lentils in Saskatchewan. 2010. Accessed on Oct 15, 2010.
87. Saskatchewan Pulse Growers, 2000 http://www.saskpulse.com/. Accessed December, 2000.
88. Schweiggert R.M., Kopec R.E., Villalobos-Gutierrez M.G., Högel J., Quesada S., Esquivel P., Schwartz S.J., Carle R., 2014. Carotenoids are more bioavailable from

papaya than from tomato and carrot in humans: A randomised cross-over study. The British Journal of Nutrition 111: 490– 98.

89. Sell R., 1993. Lentil. NDSU Extension service. Accessed November 7, 2010.

90. Singh M., Bisht I.S., Dutta M., Kumar K., Kumar S., Bansal K.C., 2014. Genetic studies on morpho-phenological trailts in lentil (*Lens culinaris* Medik). Journal of Genetics 93: 561– 66.

91. Sharma B., Emami M.K., 2002. Discovery of a new gene causing dark green cotyledons and pathway of pigment synthesis in lentil (*Lens culinaris* Medik). Euphytica 124: 349– 53.

92. Sharpe A.G., Ramsay L., Sanderson L.A., Fedoruk M.J., Clarke W.E., Li R., Kagale S., Vijayan P., Vandenberg A., Bett K.E., 2013. Ancient orphan crop joins modern era: Gene-based SNP discovery and mapping in lentil. BMC genomics 14: 192.

93. Sherry C.L., Oliver J.S., Renzi L.M., Marriage B.J., 2014. Lutein supplementation increases breast milk and plasma lutein concentrations in lactating women and in infant plasma concentrations but does not affect other carotenoids. Journal of Nutrition 144: 1256– 63.

94. Siong T.E., Heng G.A., Choo K.S., 1995. Carotenoid composition and content of legumes, tubers and starchy roots by HPLC. Malaysian Journal of Nutrition 1: 51– 61.

95. Smith, J.D. and Kinman, M.L., 1965. The use of parent-offspring regression as an estimator of heritability. Crop Science 5: 595– 596.

96. Slinkard A.E., 1978. Inheritance of cotyledon color in lentils. Journal of Heredity 69: 139– 40.

97. Stahl W., Sies H., 1996. Lycopene: A biologically important carotenoid for humans? Archives of Biochemistry and Biophysics 336: 1– 9.

98. Takaichi S., 2014. General methods for identification of carotenoids. Biotechnology Letters: 1127– 28.

99. Tee E., Goh A.H., Khor S.C., 1995. Carotenoid composition and content of legumes, tubers and starchy roots by HPLC. Malaysian Journal of Nutrition 1: 63– 74.

100. Thavarajah D., Thavarajah P., Sarker A., Vandenberg A., 2009. Lentils (*Lens culinaris* Medikus subspecies c*ulinaris*): A whole food for increased iron and zinc intake. Journal of Agricultural and Food Chemistry 57: 5413– 19.

101. Tian L., Musetti V., Kim J., Magallanes-Lundback M., DellaPenna D., 2004. The *Arabidopsis LUT 1* encodes a member of the cytochrome P450 family that is required for carotenoid ε-ring hydrocylation activity. Proceeding of the National Academy of Science of the United States of America 101: 402– 407.

102. Wang N., Daun J.K., 2004. Effect of variety and crude protein content on nutrients and certain antinutrients in field (*Pisum sativum*). Journal of the Science of Food and Agriculture 84: 1021– 29.

103. Wang N., Daun J.K., 2006. Effects of variety and crude protein content on nutrients and anti-nutrients in lentils (*Lens culinaris*). Food Chemistry 95: 493– 502.

104. Warkentin T., Kaliyaperumal A., Ramachandran A.,Thavarajah D., Tar'an, B., Bett, K.E., 2010. First annual report to agriculture development fund and Saskatchewan pulse growers. Crop Development Center, University of Saskatchewan.

105. Young A.J., Lowe G.M., 2000. Antioxidant and peroxidant properties of carotenoids. Archives of Biochemistry and Biophysics 385: 20– 27.

106. Zhou C.,Han L., Pislariu C.,Nakashima J., Fu C., Jiang Q., Quan L., Blancaflor E.B., Tang Y., Bouton J.H., Udvardi M., Xia G., Wang Z., 2011. From model to crop: functional analysis of a stay-green gene in the model legume *Medicago truncatula* and effective use of the gene for alfalfa improvement. Plant Physiology 157: 1483– 1496.

107. Zhu C., Gore M., Buckler E. S., Yu J., 2008. Status and prospects of association mapping in plants. The Plant Genome 1: 5– 20.

Appendices

Appendix 1 Observed offspring segregation ratio, value of X^2 test for yellow, red and green cotyledon colour among F_3 generations in each cross

	$F_{2:3}$ population for cross 6116				$F_{2:3}$ population for cross 6124					$F_{2:3}$ population for cross 6126			
Selected F_2 plant	Yellow	Green	Total	Selected F_2 plant	Red	Yellow	Green	Total	Selected F_2 plant	Red	Yellow	Total
6116s-4	25	-	25	6124s-5	16	8	-	29	6126s-5	-	24	24
6116s-16	20	9	29	6124s-10	-	15	9	24	6126s-8	-	29	29
6116s-20	22	12	34	6124s-14	20	9	12	41	6126s-16	9	7	16
6116s-26	9	6	15	6124s-19	-	-	35	35	6126s-27	19	9	28
6116s-55	-	24	24	6124s-25	16	7	11	34	6126s-39	-	33	33
6116s-60	15	4	19	6124s-30	15	7	9	31	6126s-51	22	5	27
6116s-66	25	12	37	6124s-44	-	-	15	15	6126s-62	12	7	19
6116s-84	28	13	41	6124s-70	16	4	7	26	6126s-67	15	-	15
6116s-95	13	4	17	6124s-93	18	5	9	32	6126s-108	15	6	21
6116s-123	-	22	22	6124s-113	-	18	8	26	6126s-110	27	6	33
	157	106	263		101	73	115	293		119	126	245

Appendix 2 Association number, accession name, country of origin and their STRUCTURE sub-group assignment.

Number	Accession	Sub-Group Assignment	Origin	Cotyledon Colour
1	ILL_0028	Cluster1	Syria	Y
2	ILL_0293	Cluster1	Greece	R
3	ILL_0313	Cluster1	Palestine	R
4	ILL_1139	Cluster1	Lebanon	R
5	ILL_2433	Cluster1	Ethiopia	R
6	ILL_4400	Cluster1	Syria	R
7	ILL_4671	Cluster1	USA	R
8	ILL_4768	Cluster1	Yemen	R
9	ILL_5151	Cluster1	India	R
10	ILL_5511	Cluster1	Syria	Y
11	ILL_5883	Cluster1	Jordan	Y
12	ILL_6853	Cluster1	Syria	Y
13	ILL_7585	Cluster1	Turkey	Y
14	PI_273664	Cluster1	Ethiopia	R
15	PI_298631	Cluster1	Peru	Y
16	PI_329169	Cluster1	Iran	R
17	PI_420929	Cluster1	Jordan	R
18	ILL_1220	Cluster2	Iran	R
19	ILL_1553	Cluster2	Iran	Y
20	ILL_2501	Cluster2	India	Y
21	ILL_2607	Cluster2	India	Y
22	ILL_3025	Cluster2	India	R
23	ILL_3347	Cluster2	India	Y
24	ILL_3597	Cluster2	India	Y
25	ILL_4164	Cluster2	India	R
26	ILL_4609	Cluster2	Netherlands	Y
27	ILL_4875	Cluster2	Uzbekistan	Y

28	ILL_6967	Cluster2	Brazil	Y
29	PI_178971	Cluster2	Turkey	Y
30	PI_251032	Cluster2	Iran	G
31	PI_300250	Cluster2	Syria	R
32	PI_308614	Cluster2	Syria	R
33	PI_426803	Cluster2	Pakistan	R
34	PI_431662	Cluster2	Iran	Y
35	PI_431679	Cluster2	Iran	R
36	PI_431705	Cluster2	Iran	R
37	PI_431710	Cluster2	Iran	R
38	PI_431714	Cluster2	Iran	R
39	PI_431717	Cluster2	Iran	Y
40	PI_431756	Cluster2	Iran	R
41	ILL_3502	Cluster3	Ukraine	Y
42	ILL_4665	Cluster3	Hungary	R
43	ILL_4740	Cluster3	France	R
44	ILL_4804	Cluster3	Libya	Y
45	ILL_4956	Cluster3	Portugal	R
46	ILL_6182	Cluster3	Tunisia	Y
47	ILL_7089	Cluster3	Russia	Y
48	ILL_7537	Cluster3	Argentina	Y
49	ILL_7747	Cluster3	Syria	Y
50	PI_178939	Cluster3	Turkey	Y
51	PI_297284	Cluster3	Argentina	Y
52	PI_298922	Cluster3	Italy	R
53	PI_299121	Cluster3	Mexico	R
54	PI_299126	Cluster3	Mexico	R
55	PI_299215	Cluster3	Chile	R
56	PI_320954	Cluster3	Hungary	R
57	PI_339292	Cluster3	Turkey	R
58	PI_343026	Cluster3	Former Soviet Union and Former Serbia	R

59	PI_357225	Cluster3	Montenegro	G
60	PI_368647	Cluster3	Macedonia	R

R- red, Y- yellow, G- green

Appendix 3 Significant markers association with corrected *P*-values for the concentration of lutein, violaxanthin and zeaxanthin using the GLM model

Number	Carotenoids	Marker	p-value
1	Lut	LcC03973p328	5.45E-05
2	Vio	LcC06688p111	5.79E-05
3	Zea	LcC07588p354	7.11E-05
4	Vio	LcC08413p299	9.26E-05
5	Vio	LcC08708p388	0.000106
6	Zea	LcC07801p998	0.000115
7	Vio	LcC07664p815	0.000124
8	Vio	LcC06318p468	0.000139
9	Vio	LcC00283p671	0.000216
10	Vio	LcC15826p559	0.000304
11	Vio	LcC01915p686	0.000333
12	Vio	LcC04047p912	0.000338
13	Vio	LcC01012p1388	0.000346
14	Zea	LcC00678p101	0.000379
15	Zea	LcC07666p460	0.00042
16	Vio	LcC08231p98	0.000466
17	Lut	LcC00678p101	0.000528
18	Zea	LcC05316p244	0.000609

19	Vio	LcC04768p149	0.000645
20	Lut	LcC03292p350	0.000645
21	Zea	LcC05912p421	0.000664
22	Lut	LcC04252p559	0.000666
23	Zea	LcC18294p329	0.000791
24	Zea	LcC04252p559	0.00082
25	Vio	LcC06739p564	0.000861
26	Vio	LcC02442p413	0.000874
27	Zea	LcC03366p499	0.000879
28	Lut	LcC24898p519	0.000888
29	Lut	LcC20366p221	0.000918
30	Vio	LcC16457p169	0.000937
31	Zea	LcC21377p693	0.000977
32	Vio	LcC10510p247	0.001
33	Zea	LcC21183p306	0.00105
34	Vio	LcC01308p107	0.00107
35	Zea	LcC11894p84	0.00111
36	Vio	LcC04942p286	0.00112
37	Vio	LcC05316p244	0.00129
38	Lut	LcC05658p269	0.00132
39	Vio	LcC03135p421	0.00132

40	Zea	LcC06625p437	0.00142
41	Zea	LcC05323p364	0.00145
42	Zea	LcC10975p380	0.00147
43	Vio	LcC00043p429	0.00149
44	Vio	LcC14588p324	0.00151
45	Zea	LcC02372p366	0.00151
46	Lut	LcC10717p585	0.00157
47	Zea	LcC05094p129	0.0016
48	Zea	LcC06883p495	0.00162
49	Zea	LcC00599p845	0.00168
50	Lut	LcC07588p354	0.00174
51	Zea	LcC27189p258	0.00174
52	Zea	LcC11344p81	0.00175
53	Zea	LcC15790p714	0.00177
54	Zea	LcC00283p671	0.00191
55	Vio	LcC07760p683	0.002
56	Vio	LcC06625p437	0.00204
57	Zea	LcC10375p155	0.00204
58	Zea	LcC06323p263	0.00206
59	Vio	LcC00628p872	0.00214
60	Zea	LcC20366p221	0.00215

61	Lut	LcC01012p1388	0.00226
62	Vio	LcC12719p848	0.00239
63	Vio	LcC00556p1386	0.00246
64	Vio	LcC23105p294	0.00252
65	Zea	LcC01908p896	0.00254
66	Vio	LcC03529p437	0.00258
67	Zea	LcC25737p350	0.0027
68	Vio	LcC01743p221	0.00271
69	Vio	LcC00092p540	0.00281
70	Vio	LcC08664p385	0.00283
71	Lut	LcC07666p460	0.00284
72	Vio	LcC07022p72	0.003
73	Zea	LcC23441p269	0.00312
74	Lut	LcC01915p686	0.00323
75	Vio	LcC07291p538	0.00323
76	Vio	LcC23441p269	0.00324
77	Vio	LcC00709p1412	0.00331
78	Vio	LcC01215p275	0.00331
79	Vio	LcC10506p467	0.00338
80	Vio	LcC07666p460	0.00345
81	Vio	LcC03136p479	0.00349

82	Zea	LcC03809p187	0.0035
83	Zea	LcC01001p523	0.00361
84	Vio	LcC06883p495	0.00363
85	Vio	LcC02612p585	0.00371
86	Lut	LcC16599p253	0.00372
87	Vio	LcC11648p105	0.00373
88	Vio	LcC05543p176	0.00383
89	Lut	LcC06883p495	0.00414
90	Vio	LcC21570p454	0.00414
91	Vio	LcC03287p157	0.00423
92	Lut	LcC01560p267	0.00428
93	Zea	LcC05440p266	0.0046
94	Vio	LcC02056p548	0.00469
95	Lut	LcC23105p294	0.00471
96	Vio	LcC01164p240	0.00475
97	Vio	LcC02234p372	0.00486
98	Zea	LcC05913p392	0.005
99	Vio	LcC04091p498	0.00507
100	Vio	LcC02468p394	0.0053
101	Vio	LcC04347p392	0.00531
102	Vio	LcC23123p268	0.00538

103	Zea	LcC14588p324	0.00541
104	Vio	LcC25737p350	0.00546
105	Vio	LcC20137p84	0.00547
106	Vio	LcC09549p442	0.00548
107	Zea	LcC01560p267	0.0055
108	Vio	LcC14401p336	0.00551
109	Vio	LcC07457p534	0.00557
110	Vio	LcC03366p499	0.00561
111	Vio	LcC03880p98	0.00575
112	Vio	LcC05912p421	0.00599
113	Vio	LcC09871p177	0.00607
114	Vio	LcC23289p212	0.00607
115	Vio	LcC04960p169	0.00625
116	Vio	LcC01206p976	0.00633
117	Vio	LcC10566p589	0.00636
118	Vio	LcC03534p135	0.00641
119	Vio	LcC06975p327	0.00642
120	Vio	LcC06600p499	0.00663
121	Zea	LcC28142p190	0.00669
122	Vio	LcC07588p354	0.007
123	Zea	LcC05901p668	0.007

124	Vio	LcC06139p91	0.00701
125	Vio	LcC01767p155	0.00703
126	Vio	LcC23157p103	0.00717
127	Lut	LcC00104p269	0.00727
128	Lut	LcC24457p390	0.00733
129	Vio	LcC07680p141	0.00744
130	Zea	LcC03973p328	0.0076
131	Lut	LcC03228p321	0.00774
132	Vio	LcC06523p522	0.00796
133	Lut	LcC01537p474	0.00808
134	Zea	LcC05942p621	0.00829
135	Zea	LcC17753p341	0.0083
136	Vio	LcC03388p265	0.00832
137	Zea	LcC18665p480	0.00849
138	Zea	LcC04599p195	0.00861
139	Lut	LcC00043p429	0.00882
140	Zea	LcC00172p513	0.00893
141	Lut	LcC17934p266	0.00894
142	Zea	LcC00805p364	0.00895
143	Vio	LcC01745p1527	0.00918
144	Vio	LcC17953p450	0.00919

145	Zea	LcC14401p336	0.00951
146	Vio	LcC17052p247	0.00952
147	Zea	LcC02330p709	0.00954
148	Lut	LcC02097p103	0.00965
149	Lut	LcC01680p1198	0.00971
150	Zea	LcC01440p87	0.00971
151	Vio	LcC05465p528	0.00984
152	Lut	LcC12929p665	0.01034
153	Vio	LcC09041p299	0.01038
154	Zea	LcC01410p217	0.0107
155	Vio	LcC19753p120	0.01071
156	Zea	LcC13147p355	0.01076
157	Vio	LcC03809p187	0.0109
158	Vio	LcC02381p482	0.01097
159	Lut	LcC02372p366	0.01107
160	Lut	LcC11340p677	0.01118
161	Vio	LcC00654p239	0.01132
162	Zea	LcC24898p519	0.01135
163	Vio	LcC06218p492	0.01139
164	Vio	LcC04356p79	0.01145
165	Vio	LcC10709p228	0.01165

166	Vio	LcC06877p157	0.0119
167	Vio	LcC21501p760	0.01194
168	Vio	LcC02323p303	0.01206

Appendix 4 STRUCTURE sub-groups and each line's assigned groupings. The level of admixture that each accession has is shown by each value within the sub-groups.

Number	Accession	STRUCTURE Sub-groups			Sub-Group Assignment
		Cluster1	Cluster2	Cluster3	
1	ILL_0028	0.897	0.002	0.101	Cluster1
2	ILL_0293	0.84	0.005	0.155	Cluster1
3	ILL_0313	0.615	0.053	0.332	Cluster1
4	ILL_1139	0.732	0.252	0.016	Cluster1
5	ILL_1220	0.247	0.467	0.287	Cluster2
6	ILL_1553	0.227	0.424	0.348	Cluster2
7	ILL_2433	0.999	0.001	0	Cluster1
8	ILL_2501	0	0.999	0	Cluster2
9	ILL_2607	0	0.999	0	Cluster2
10	ILL_3025	0	0.999	0	Cluster2
11	ILL_3347	0	0.999	0	Cluster2
12	ILL_3502	0.288	0.086	0.626	Cluster3
13	ILL_3597	0	0.999	0	Cluster2
14	ILL_4164	0.200	0.559	0.241	Cluster2
15	ILL_4400	0.928	0.001	0.072	Cluster1
16	ILL_4609	0.001	0.999	0	Cluster2
17	ILL_4665	0.384	0.001	0.615	Cluster3
18	ILL_4671	0.788	0	0.212	Cluster1
19	ILL_4740	0	0	0.999	Cluster3
20	ILL_4768	0.999	0	0	Cluster1
21	ILL_4804	0.218	0.106	0.675	Cluster3
22	ILL_4875	0.07	0.906	0.024	Cluster2
23	ILL_4956	0.067	0.004	0.929	Cluster3
24	ILL_5151	0.997	0.003	0	Cluster1
25	ILL_5511	0.719	0	0.281	Cluster1

26	ILL_5883	1	0	0	Cluster1
27	ILL_6182	0.141	0.116	0.743	Cluster3
28	ILL_6853	0.869	0.002	0.13	Cluster1
29	ILL_6967	0.410	0.419	0.171	Cluster2
30	ILL_7089	0.150	0.001	0.848	Cluster3
31	ILL_7537	0.470	0.002	0.529	Cluster3
32	ILL_7585	0.47	0.311	0.218	Cluster1
33	ILL_7747	0	0.001	0.999	Cluster3
34	PI_178939	0	0	0.999	Cluster3
35	PI_178971	0.001	0.998	0.002	Cluster2
36	PI_251032	0.082	0.624	0.293	Cluster2
37	PI_273664	0.999	0	0	Cluster1
38	PI_297284	0.349	0	0.65	Cluster3
39	PI_298631	0.819	0.003	0.178	Cluster1
40	PI_298922	0.121	0.055	0.824	Cluster3
41	PI_299121	0.282	0.002	0.716	Cluster3
42	PI_299126	0.176	0.011	0.813	Cluster3
43	PI_299215	0.002	0.472	0.526	Cluster3
44	PI_300250	0.202	0.582	0.216	Cluster2
45	PI_308614	0.183	0.593	0.224	Cluster2
46	PI_320954	0	0	0.999	Cluster3
47	PI_329169	0.514	0.276	0.21	Cluster1
48	PI_339292	0.137	0.081	0.782	Cluster3
49	PI_343026	0.079	0.436	0.485	Cluster3
50	PI_357225	0.001	0	0.999	Cluster3
51	PI_368647	0	0.001	0.999	Cluster3
52	PI_420929	0.999	0	0	Cluster1
53	PI_426803	0.002	0.997	0.002	Cluster2
54	PI_431662	0	1	0	Cluster2
55	PI_431679	0	0.999	0	Cluster2
56	PI_431705	0	0.999	0	Cluster2

57	PI_431710	0	0.999	0	Cluster2
58	PI_431714	0.001	0.999	0	Cluster2
59	PI_431717	0	0.999	0	Cluster2
60	PI_431756	0	0.999	0	Cluster2

Appendices 5 Carotenoid concentration calculation

A=e l c

Calculation for lutein for example

Where, A = .75, this was multiplied by 10 was the depth of the plate was 10 mm

E (extinction coefficient) = 145000 (http://epic.awi.de/28865/1/Jef1997au.pdf)

And the length of each cell in the plate for UV wavelength to pass was 1mm

We have to find the concentration of the standard

So, 7.5 = 145000 *1* C

C= 7.5/ 145000 = 0.000051724 mol cm

= 0.051724 millimol m

We know that the molecular weight of Lutein is = .568.88 g/mol (http://epic.awi.de/28865/1/Jef1997au.pdf)

= .56888 milli g/ mol

1 mol = 0.029629 milli g/mol (amount injected into the HPLC i.e per 20 µl) (1/0.51724 = 19.3 times)

Concentration in 1000 µl, 1000/20 = 50

0.029629/50 = 0.0000592583 milli g

= 0.0592583 µg